JOURNAL OF SPACE MISSION ARCHITECTURE

JSMA

Issue 1 Fall 1999

A PUBLICATION OF THE JET PROPULSION LABORATORY'S CENTER FOR SPACE MISSION ARCHITECTURE AND DESIGN, A NASA CENTER OF EXCELLENCE

NOTICE TO AUTHORS

The Journal of Space Mission Architecture is published by the Jet Propulsion Laboratory's Center for Space Mission Architecture and Design. The Journal's goal is to provide a medium for the presentation of space mission architecture concepts, processes and tools. Authors are invited to submit articles on all areas of space mission architecture, such as: case studies of the development of missions including new or unusual architectural challenges, non-NASA mission architectures, the mission architecture process and tools, historical or biographical notes, and space program architecture.

Authors are referred to:

http://www.jpl.nasa.gov/csmad/journal/

for instructions on the submission of articles.

All correspondence should be addressed to the Editor-in-Chief:

DONNA M. WOLFF
Jet Propulsion Laboratory
301-170U
4800 Oak Grove Drive
Pasadena CA 91109

This publication was prepared by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Foreword

As the word is applied to the design of buildings, cities and the like, architecting is a very old craft. The need for an individual to maintain a sense of what a design is *for* in addition to how it is made, is an ancient notion. In all senses of the profession, architects pursue an integration of function and form as they translate goals into reality. Whether the architect's job involves the design of a building or another type of system, he/she initially considers the objectives (or goals), possible solutions, and the realities of cost and schedule in order to lay down a basis on which more details can be hung. Then, as the design progresses from concept into reality, the architect continues to ensure that the progression remains true to the initial dream.

The craft of *systems* architecting develops systems solutions to the design of a complex product. Taken literally, the term does apply to the more traditional building or city task — these are certainly good examples of complex systems — but the term is generally applied to some product spanning mechanical, electronic and software disciplines. In either domain, the need for such an individual becomes pronounced as whatever is being designed becomes more complex and multi-disciplinary. By contrast, even in single disciplines (such as software), architects have been found valuable and have produced breakthrough changes in these disciplines.

To get to the immediate point, space mission architecting has now entered the picture and come into its own, as the craft has with other fields. Whereas in the past spacecraft were logically (if not physically) simple enough that architecting was the first step of the System Engineer, this field has now matured to the point where one skill is required to be concerned with the relationship between goals and requirements, quite apart from another skill concerned with the relationship between requirements and their implementation. Thus we recognize the Space Mission Architect and his/her architecture as the precursor, initial input and baseline for the System Engineer and the accompanying system design.

We dedicate this Journal to the work of the architecting role. We intend that it always be a forum where both the role of architect and the product, architecture, can be displayed, discussed and debated openly, limited only by the quality controls of peer review and readership demands. We hope that its readers find it useful, informative and, at least occasionally, entertaining.

Steve Wall
August 1999

The Mars Surveyor Program Architecture

James F. Jordan and Sylvia L. Miller

Abstract

The architecture of NASA's Mars Surveyor Program received intense scrutiny in 1998. The results of that effort are presented here. After a review of the ongoing and near-term missions to Mars, the missions in the proposed architecture for launches in 2003 and beyond are described. The heart of the new program architecture consists of missions which will return samples of Martian rock and soil back to Earth for analysis. A primary scientific goal is to understand Mars as a possible abode of past or present life. Other key elements are Mars Micromissions, Telecommunications, and steps toward human exploration. Potential international partnering is key.

I. INTRODUCTION. The architecture of NASA's program of robotic Mars exploration missions received an intense scrutiny during the summer months of 1998. We present here the results of that scrutiny, and describe a list of Mars exploration missions which are now being proposed by the nation's space agency. No decision on the final mission architectures will be made until after each mission has complied with the National Environmental Policy Act and completed an Environmental Assessment or Environmental Impact Statement.

The heart of the new program architecture consists of missions which will return samples of Martian rocks and soil back to Earth for analysis. A primary scientific goal for these missions is to understand Mars as a possible abode of past or present life. The current level of sophistication for detecting markers of biological processes and fossil or extant life forms is much higher in Earth-based laboratories than possible with remotely deployed instrumentation, and will remain so for at least the next decade. Hence, bringing Martian samples back to Earth is considered the best way to search for the desired evidence.

A Mars sample return mission takes approximately three years to complete, as seen in Figure 1. Transit from Earth to Mars requires most of a year. After a lapse of about a year at Mars, during which time orbital and surface operations can take place, and the correct return launch energy constraints are met, a Mars-to-Earth return flight can be initiated. This return leg also takes approximately one year. Opportunities to launch these 3-year sample return missions occur about every 2 years, as shown in Figure 2. The figure depicts schedules for possible flights to and from Mars for Earth launches in 2003, 2005, 2007 and 2009. Transits for less than a 180° flight angle ("Type 1"), measured from the sun, and more than 180° ("Type 2"), are both shown.

II. BACKGROUND MISSIONS. The current and near-term planned missions, which are predecessors to future sample return missions, are briefly described next. They are summarized in Figure 3.

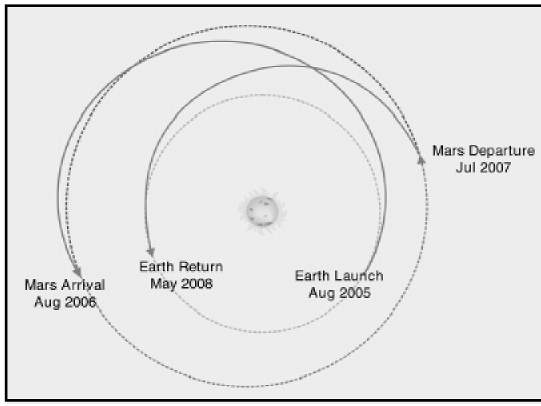


Figure 1. A sample return mission takes approximately three years to complete. Example shown is the 2005 opportunity.

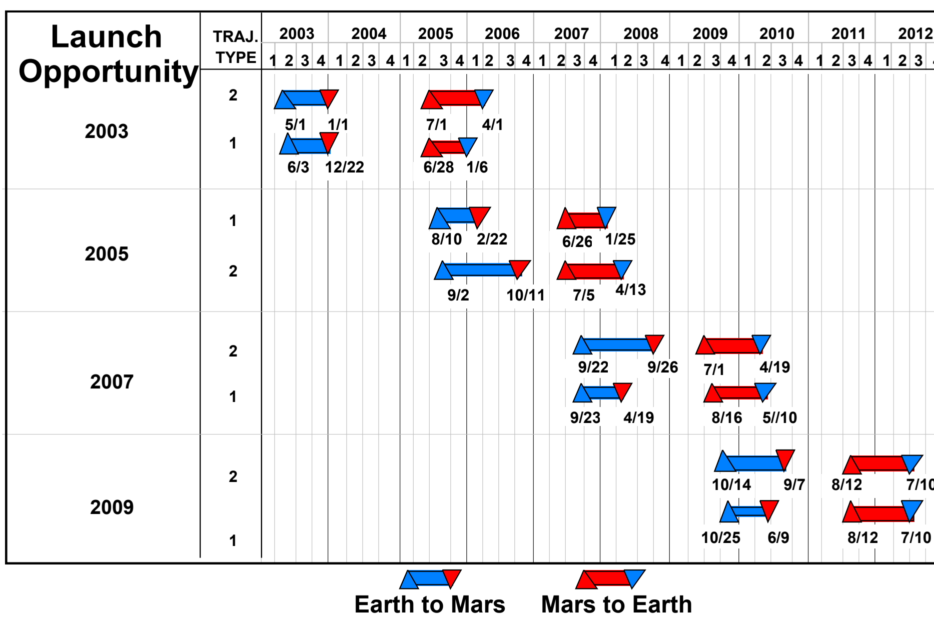


Figure 2. Key Mars opportunities for launch/return.

In the 1996 launch opportunity, NASA sent two missions to Mars and thus began its first return to the red planet in over two decades. One of these missions was Mars Pathfinder, part of NASA's Discovery Program. Pathfinder consisted of a lander and a rover named Sojourner that was about the size of a microwave appliance. They landed on July 4, 1997, and gathered significant scientific data on the soil, rocks, and other elements of their environment and demonstrated important new technologies. In addition, Mars Pathfinder generated unprecedented public interest, as demonstrated by the record-breaking activity on the World Wide Web.

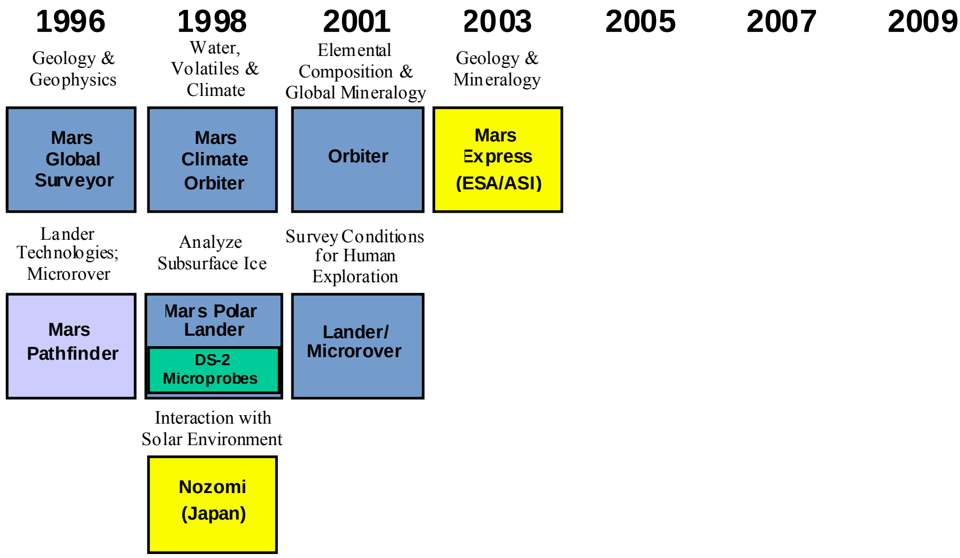


Figure 3. Mars exploration background missions.

The other mission launched in 1996, an orbiter called Mars Global Surveyor (MGS), is the first in a series of missions comprising NASA's Mars Surveyor Program. The Mars Surveyor Program was formally established in 1994 to embark on a continuous exploration of Mars with the long-range goal of understanding Mars in terms of life, climate and resources (see Figure 4). The missions are to look for evidence of past or present life; to understand the weather, climatic processes and climate history; and to identify the main environmental resources and their potential uses. A unifying theme is water — understanding where on Mars it may be now and the role it played in Martian history. The budget for the Program was initiated at about $160 M per year, including launch vehicles and mission operations. Beginning with fiscal year '98, the Program's budget was augmented by $40 to $60 M per year, along with an increased focus on the search for evidence of life. The Mars Surveyor Program is the core of the currently funded exploration of Mars.

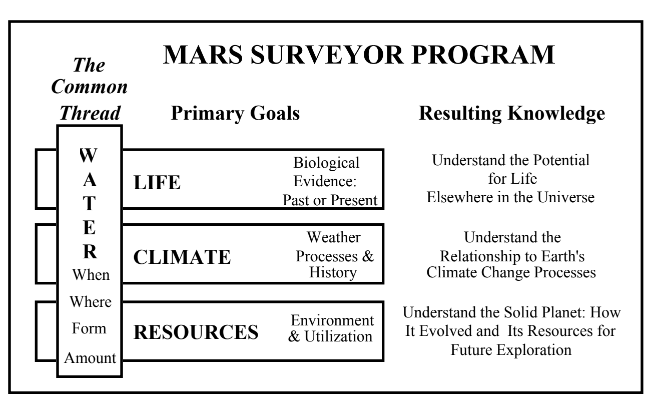


Figure 4. The primary goals of the Mars Surveyor Program.

MGS reached Mars in September 1997 and used a technique called aerobraking to reach its proper mapping orbit. This technique required that the orbiter skim the atmosphere of Mars to reduce speed and, thus, the size of its orbit. The use of the Martian atmosphere to adjust the orbit of the spacecraft greatly reduced the propellant requirement, thus enabling the use of a lower-cost launch vehicle. Even though the prime science mission just began in March 1999, a large amount of science data has already been collected, with key results relating to the planet's magnetic field, its mineralogy and topography, the topology of the northern polar cap, and thick dust on the moon Phobos.

Two spacecraft currently en route to Mars will be joining MGS toward the end of 1999. Mars Climate Orbiter, launched in December 1998, and Mars Polar Lander, launched in January 1999, will focus on the climate of Mars and search for clues to the reasons for its dramatic changes over time. Key objectives are to understand how water and dust move about the planet and to find clues to the location of water on Mars. The orbiter carries instruments to map the planet's surface, profile the structure of the atmosphere, and possibly detect surface ice reservoirs. The lander will search for evidence of water beneath the surface at its landing site near the south pole of Mars. With a robotic arm and attached camera it will dig a trench several centimeters deep and examine in detail the fine-scale layering, if any, along the walls of the trench. It will also deliver a small sample of soil to a miniature oven. An instrument will analyze the volatiles that are generated when the sample is heated and measure their concentrations. Water vapor would demonstrate that water in some form was embedded in the soil.

Mars Polar Lander is also carrying to Mars two basketball-sized probes called Deep Space 2. These probes are part of NASA's New Millennium Program, the purpose of which is to test new technologies. The microprobes will separate from the lander just before reaching Mars and will smash into the surface with a deceleration of about 80,000 g. Each probe is designed to break into two parts at this time, a penetrator and a relayer. The aftbody will remain on the surface and relay data from the probe to MGS for transmission back to Earth. The forebody will bury itself below the surface up to about a meter in depth. As does Mars Polar Lander, these penetrators also carry instruments to search, in two additional sites, for evidence of water.

The Mars Surveyor Program will continue with the launch of an additional pair of spacecraft in 2001. A lander will carry a small rover called Marie Curie, a slightly-modified flight spare of Sojourner. It will analyze the rocks and soil in a near-equatorial region to be chosen before launch, with the selected site to be targeted in flight. The lander will also carry three instruments in support of the possible future human exploration of Mars: a dust and soil characterization experiment, a demonstration of extracting oxygen from the Martian atmosphere, and a radiation monitor. The second spacecraft is an orbiter which, from a near-polar orbit, will use remote sensing instruments to map the temperatures and elemental composition of the surface, and to search for near-surface reservoirs of water. This spacecraft will also monitor radiation, but from orbit, and these measurements will be used in conjuction with those gathered on the surface to enhance significantly our understanding of the radiation environment at Mars.

The European Space Agency (ESA) and the Italian space agency (ASI) together are planning an orbiter to be launched to Mars in June 2003. The mission, called Mars Express, will be the first of a new class of “flexible” missions in the revised ESA long-term scientific program. The payload will consist of a large set of remote-sensing instruments, nearly all with significant heritage from European instruments lost with the Russian Mars 96 mission. Examples are: a stereo imager that will provide global 10-meter-resolution photogeology of the Martian surface and an infrared mapping spectrometer to obtain global mineralogy maps with 100-meter resolution. A new instrument, to be provided by an ASI/NASA partnership, is a radar sounder that will map the subsurface structure of Mars at the kilometer scale down to the permafrost. Mars Express is also planning to carry a small surface lander, called Beagle 2, provided by a UK consortium led by the Planetary Sciences Research Institute, which will land as Pathfinder did, using airbags. It will carry a 60-kg package of exobiology, geochemistry, and atmospheric chemistry investigations.

A Japanese mission called Nozomi was launched on July 4, 1998, exactly one year after the Pathfinder landing. Although originally scheduled to arrive at Mars in 1999, it will actually reach Mars four years later. From a highly elliptical orbit, Nozomi will study the structure and dynamics of the upper atmosphere of Mars and its interaction with the solar wind.

Arriving in 2006 will be NetLander, a network of 4 small spatially dispersed surface stations which will make simultaneous measurements on magnetism, seismology, and meteorology. To be built in Europe, these landers, each within its own aeroshell, will be carried to Mars by the CNES orbiter being planned for launch in 2005 (see below).

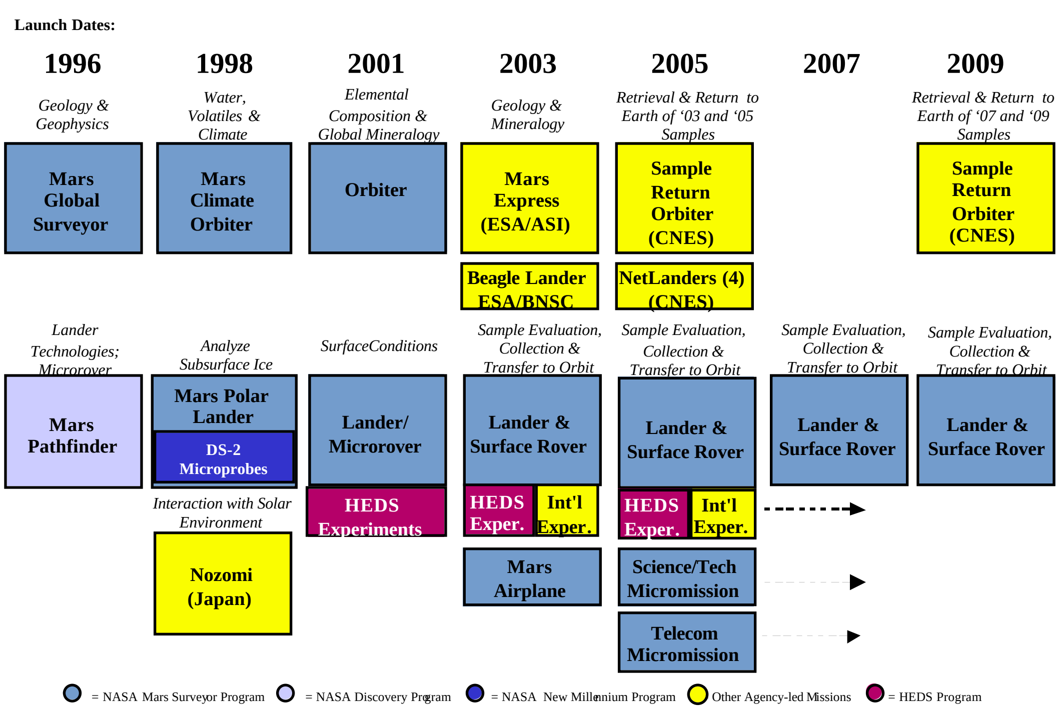


Figure 5. Missions in the Mars exploration architecture.

III. SAMPLE RETURN MISSIONS. As depicted in Figure 5, the sample return mission architecture begins in 2003 with the launch of a large Martian surface lander on a Delta III or Atlas IIIA class launch vehicle. The lander will carry a large, 75-kg surface rover which will be deployed to cover a wide area, perhaps a square kilometer or more, to search for suitable rocks to sample. The rover will use a variety of instruments to select the most promising rocks which might contain biological evidence. Rover instrumentation will drill into several selected rocks, obtain small sample cores from each, and return the sets to the lander. The rover may perform multiple excursions to collect rock and soil samples. The rover will deposit these samples in the payload canister of a solid-rocket-propelled ascent vehicle. The lander will carry a coring drill which will allow subsurface samples to augment those collected by the rover for a total sample mass of about .5 kg. The ascent rocket will be fired to lift the payload sample canister into a circular orbit, 600 km above the Martian surface, where it will await pickup by an orbiter.

In 2005 NASA is planning to send a second lander/rover combination to Mars on a French Ariane 5 launch vehicle. The 2005 lander and rover will be duplicates of the 2003 vehicles, and perform essentially the same functions — depositing a second sample-filled canister into a nearly identical Martian orbit.

The same Ariane 5 launch vehicle which carries the U.S. lander will also carry a Mars-orbiting spacecraft supplied by the French national space agency (CNES). This Mars orbiter will pick up both the canisters and bring them back to Earth. The orbiter is first maneuvered into the same orbit as the canisters and then will autonomously rendezvous and dock with each canister, in turn, using radio signals and laser radar sightings. Once collected by the spacecraft, the canisters will be imbedded in Earth-return atmospheric re-entry capsules. The capsules are carried back to Earth by the orbiter and released into the Earth's atmosphere in 2008, after which they will land at a suitable retrieval site. Figure 6 depicts a schematic of the 2003 and 2005 missions.

By implementing this baseline plan, the missions described above will bring two distinct sample sets back to Earth in 2008 from two distinct locations on Mars. These highly protected samples will then be analyzed in Earth laboratories with a focus on the search for markers of biological processes or evidence of fossil or existing life forms.

IV. MICROMISSIONS. The sample return missions as described above are expected to comprise the backbone of the NASA Mars exploration program for the 2003–2005 launch opportunities. Other major missions, perhaps more sample return missions, will follow. In addition to the major mission undertakings, NASA also plans a line of smaller complementary missions, called micromissions. These micromissions will gather knowledge only, not samples for return, and will continue the NASA solar system exploration tradition of global planetary surveying and selected, high-resolution site reconnaissance.

The micromission line will begin with launches in 2003 or 2005, depending on the available budget wedge. A generic spacecraft bus will be developed that can be replicated to carry science payloads to Mars for several opportunities. These micromissons may be launched on small, U.S.-supplied launchers like the Taurus, or as auxiliary payloads on commercial flights of the French Ariane vehicle. Science payload mass will approach 20 kg on the surface of Mars or 10 kg in orbit about Mars. Micromisson payloads may include surface penetrators, aerial platforms like atmospheric balloons and gliders, small landers and remote-sensing orbiters. Micromissions may also deliver communications satellites to Mars orbit (see Section VII).

V. INTERNATIONAL CONTRIBUTIONS. NASA thus envisions a balance of missions in the Mars program architecture, with major sample return endeavors both within the U. S. and in partnership with CNES, and complimentary science data gathering micromissons whose carrier bus is developed by American industry. Below we describe some of the expected contributions from the international partners.

Space agencies around the world have a strong interest in Mars exploration. This interest, along with NASA's plans for a challenging program capped by a modest budget, creates a natural potential for international partnerships. Pursuing such relationships has been a guideline for the Mars Surveyor Program from its beginning. Discussions have been underway with a number of space agencies for several years to explore and develop opportunities for cooperation. Although formal agreements have not yet been completed, current assumptions for the architecture are described here. Additional cooperative endeavors are also possible.

To establish the framework for the discussion, the key elements envisioned for NASA's contribution are described first.

NASA will provide the landers, sample-gathering rovers and Mars ascent vehicles beginning with the 2003 launch. They may also provide some additional lander-based science payload starting with the 2005 launch. On the orbiters, to be provided by the French (see Figure 6), NASA will supply the equipment needed to find and track the sample canisters. This will consist of a receiver to track the beacons mounted on the orbiting canisters. In addition, on each orbiter NASA will provide the sample capturing device, yet to be defined, as well as the Earth entry capsules into which the samples will be inserted and eventually ride safely through Earth's atmosphere to the landing site.

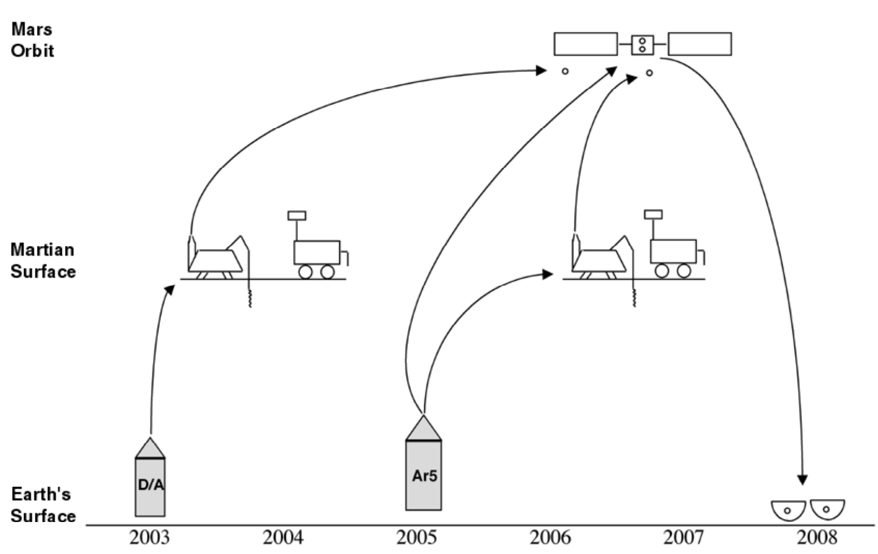


Figure 6. Schematic of Mars 2003/2005 sample return missions.

NASA has several contributions to Mars Express (see earlier discussion). Through the Mars Surveyor Program it has a partnership on the Italian-led radar sounder and supports numerous co-investigators on other experiments. Through the Discovery Program, it provides an instrument to analyze space plasmas and energetic atoms.

Other NASA contributions include launch vehicles in the class of Delta III/Atlas IIIA or Delta IV/Atlas V, beginning in 2003 and continuing as needed, and the design and production of the micromission bus, as well as some associated payloads.

The French Space Agency, CNES, will fill a key role in the baseline architecture for the Mars Surveyor Program. According to this plan, one essential item that CNES will provide is an Ariane 5 launch vehicle for the 2005 opportunity (only). Launching from French Guyana, this European rocket will have the capability to inject a 5200-kg payload to Mars at the required launch energy of 18 km^2/s^2. This is sufficient to carry both the orbiter and lander which, after separating from the launch structure, will cruise independently to Mars.

A second major contribution by CNES is the sample return orbiter, which will be launched starting in 2005. After entering Mars orbit, employing an aerocapture technique for the first time, the 2005 orbiter will have enough propellant to match orbits with both the 2003 and 2005 sample canisters in turn. Although the initial stages of each rendezvous will be directed through ground interaction (using data from the U.S.-supplied tracking equipment carried on the orbiter), the final phase, including capture of the canister, will necessarily be autonomous. The orbiter will inject out of Mars orbit, bring the Earth entry capsules back to Earth, and accurately target them for touching down at the designated landing site. After releasing the capsules, the orbiter will perform a deflection maneuver and fly past the Earth. In addition to the sample return function, the orbiter will deliver NetLander to Mars. The four probes will be released prior to Mars orbit insertion.

CNES will also provide the launch opportunities for micromissions on commercial Ariane vehicles, perhaps as early as 2003, and contribute to some payloads.

ASI, the Italian space agency, will provide a drill for the landers launched in 2003 and beyond to enable subsurface samples to be collected and included in the sample canister, along with the rover-based sample set. This complementary sample type, perhaps with an additional Italian science package, will add important diversity of science data and robustness to the missions.

ASI has a major role in the Mars Express orbiter in that they provide the telecom system. This system provides two-way communication between the spacecraft and Earth and also provides the telecom relay function for the Beagle 2 lander. It may also be used to provide this same relay function for the NetLanders. Because of the existence of this telecom capability in the 2004 to 2007 time frame, it could be applied to enhance the communication between Mars Surveyor-landed elements and Earth. It could also help with the task of locating and positioning the first sample canister in Mars orbit in 2004, providing valuable a priori data for operating the Mars Surveyor 2005 orbiter. In addition to providing the telecom function, ASI will provide some members of the operations team for Mars Express. This budget sharing with ESA will enable the operational lifetime to be extended into mid-2007. In addition, one of the Earth-based stations that may be used by Mars Express is the Italian antenna being developed on Sardinia. Among the ASI roles in the Mars Express payload, the Principal Investigator for the sub-surface radar sounder is Italian, as previously mentioned.

ESA is responsible for the Mars Express mission. They will provide the launcher, orbiter, and operations team, the last jointly with ASI. The orbiter will carry a robust payload and gather important science data, as described early in this paper, data which will complement and enhance the Mars Surveyor Program in many areas. In particular, this orbiter is the vehicle for delivering and operating the ASI/NASA subsurface radar sounder at Mars. In addition, Mars Express may provide to the Mars Surveyor operations team a priori data on the position of the 2003 sample canister once it has reached Mars orbit. The 2005 sample return orbiter could then be targeted for an orbit insertion that makes the rendezvous with the sample canisters more efficient, both in terms of schedule and propellant.

VI. LONG-TERM PROGRAM. Several options exist for a continuation of the program architecture beyond the 2003 and 2005 sample return missions. If desired, the sample return missions launched in 2003 and 2005 can be repeated in 2007 and 2009 and could produce a return of four sample sets back to Earth from four distinct Martian sites by 2012. However, many alternatives will be studied in the coming months. The possible infusion of new technologies could have important implications for these more distant missions. If sampling is continued, these technologies could lead to increases in the amount and diversity of the returned sample sets as well as increases in the area of collection coverage.

The pace of sample returns can be slowed down in future years, with the reclaimed budget deployed to increase the number of micromission per launch opportunity or to refocus the direction of the program.

VII. TELECOMMUNICATIONS. Primary communications with the 2003- and 2005-launched landers on Mars will be achieved through an X-band radio link between the landers and the NASA Deep Space Network. In addition, the 2003 and 2005 rovers will be communications-compatible with other orbiting assets like the NASA 2001 Surveyor orbiter and the ESA Mars Express orbiter. Also, some of the micromissions, described earlier, to be launched to Mars as early as 2003, are expected to be dedicated communications orbiters. Each one would service the landed elements for at least five years. The first would likely be placed into a low equatorial orbit, with later spacecraft in the network placed at various other inclinations. Current plans also call for aerostationery orbiters to be launched, one perhaps as early as 2007. Thus, the Mars Surveyor Program will be supplying a permanent telecommunication infrastructure at Mars which will enhance the performance of the science missions as well as return higher data rate information back to Earth. With these higher data rates, the story of Martian exploration may be available in real-time around the world.

VIII. HUMAN EXPLORATION. Ultimately, NASA expects that the character of the Martian exploration program will shift from purely robotic missions to the deployment of missions with human travelers. As we approach this new era, we will deploy experiments in the robotic program designed to characterize the Martian environment for human habitation and test those technologies needed to support human travel to and from Mars.

This is already the case with the Mars 2001 lander, which will carry experiments to measure the radiation environment on the Martian surface, test the soil and dust toxicity, and extract oxygen from the CO_2-dominated Martian atmosphere.

It is anticipated that the landers deployed for sample return in 2003 and beyond will also carry experiments which will pave the road for human travel to Mars. Potential experiments with the 2003 lander include an in-situ propellant production facility, which will manufacture and ignite a fuel suitable for human cargo lift-offs from the Martian surface. It is likely that experiments like these will continue to be deployed in the missions after 2003. In fact, NASA is studying the concept of robotic outposts on Mars, a rather substantive transitional phase to human exploration.

Summary

Mars continues to be interesting, both scientifically and culturally, to people around the globe. NASA's robotic program for continuing Mars exploration will probe the planet's secrets, with a reach for the existence of a biological past or presence through a campaign of bringing back to Earth carefully selected samples of Martian soil and rock. Information return from small missions augment the sample return main mission set. The program is very international in character and seeks to expand our knowledge of the planet itself and to test technologies required for human travel to Mars in the next century.

Acknowledgments. The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

FRANK JORDAN has a Ph.D. in Engineering Mechanics from the University of Texas at Austin. He has been a member of the technical staff at JPL since 1966. During his career at JPL, he has served as Supervisor of Precision Orbit Determination Group, Manager of the Navigation Systems Section and Manager of the Systems Division. He currently leads the Program Architecture and System Engineering Office in JPL's Space and Earth Science Programs Directorate. Jordan has received the Magellanic Premium from the American Philosophical Society, the Bronze Medal from the Royal Institute of Navigation, the Weems Award from the Institute of Navigation and two NASA Awards.

SYLVIA MILLER has worked on advanced planning and architecture for future Mars missions since 1995. She joined the Jet Propulsion Laboratory (JPL) in 1968 after receiving her bachelor's degree in mathematics from Douglass College, Rutgers University. She earned her master's degree in systems engineering in 1971 from West Coast University. During her career at JPL, Ms. Miller served as a team chief on the Infrared Astronomical Satellite mission, the Mission Design Manager on the Comet Rendezvous Asteroid Flyby project, a technical group supervisor in the Mission Design Section, and Manager of the Small Bodies Advanced Studies Office. She is currently a member of the Program Architecture and System Engineering Office. Ms. Miller received a NASA Exceptional Service Medal and is a Fellow of the British Interplanetary Society.

Experimental Results of LightSAR Mission Planning Using a Market-Based System

Randii R. Wessen, David Porter, Jeffrey Hilland

Abstract

The allocation of scarce spacecraft resources to multiple users has always been a difficult process. This difficulty arises from the fact that there are never enough resources (e.g., data volume, integration time, spacecraft power, etc.) to meet the stated requirements of the scientific investigators, who compete with one another to acquire their desired data sets, or 'takes.' To help solve this problem, a market-based process was developed to allocate on-orbit resources for the Lightweight Synthetic Aperture Radar (LightSAR) mission, a joint NASA/commercial endeavor. LightSAR chose to evaluate the utility of a market-based system as part of its mission concept study phase. This type of system was selected based on its prior successes in allocating resources for the Cassini spacecraft and the Space Shuttle, demonstrating that a market-based system could provide results comparable to other methods currently used for allocating resources, while requiring a smaller workforce and a shorter development period.

I. INTRODUCTION. One of the most time-consuming activities performed during mission operations is the conflict-resolution process for determining which Investigator's data requests take precedence over another's. Such conflicts must be resolved any time there are multiple science objectives for a given instrument or multiple instruments with unique objectives, i.e., when the demands for spacecraft resources outstrip the available supply.

In November 1997, the LightSAR (Lightweight Synthetic Aperture Radar) Pre-Project undertook an investigation to determine whether a market-based system could be a cost-effective planning tool. Data acquisition planning for the LightSAR radar will be complex because of the anticipated large demand for use of the payload and the complicated joint NASA/commercial organization of the project.

Typically, Project officials are placed in the difficult position of not being able to give Science Investigators all the resources (e.g., payload data acquisition time) they request. Knowing this *a priori* tends to give Investigators the incentive to request more resources than they actually need. The usual process for scheduling data acquisition requests on Earth-orbiting radar spacecraft is a "committee-driven" approach. This approach requires individual Investigators to submit requests for specific spacecraft resources to a neutral party, namely the Sequence Integrators. These individuals integrate the requests into a single, time-ordered listing of events that do not violate resource constraints. The Integrator's goal is to produce a conflict-free listing that maximizes the overall return for the mission while being fair to each Investigator. In this context, "fair" means that every attempt is made to integrate each Investigator's highest ranked requests into the listing.

Sometimes referred to as a "Serial Draft" or "Serial Dictator" method, the Integrator starts with one Investigator and selects his or her highest ranked request. The Integrator then moves to the next Investigator, selecting that person's highest ranked request, until the highest ranked request from each Investigator is incorporated into the time-ordered listing. Next, the Integrator selects each Investigator's second highest ranked request. This time, however, the order is reversed — the Investigator whose highest ranked request was selected last gets their second-highest ranked request selected first, and so on. If there are not enough resources left for a request, the Integrator tries the next-highest ranked request from that Investigator's prioritized list. This selection process continues until either all of the requests are implemented or the remaining spacecraft resources cannot accommodate any other requests.

Once the Integrator develops a listing, it is presented to the Science Investigators for evaluation and comment. Typically, those Investigators whose requests were incorporated evaluate the listing quite high. Investigators whose requests did not make the list tend to evaluate the listing low. Since no direct mechanism exists to control the number of appeals, most Investigators appeal for more data acquisition time.

The appeals process involves presenting the merits of one Investigator's request over another to the Project Scientist or some other governing board, producing what is commonly refereed to as a "Dead Weight Loss." That is, any Investigator not awarded additional acquisition time has invested their time and effort but has received no return. The end result is that multiple meetings with multiple appeals and re-integrations occur until the time allocated for developing the time-ordered listing has expired.

To reduce the amount of time and workforce needed to produce a conflict-free, time-ordered listing, LightSAR decided to evaluate a market-based system. These systems have been used for centuries in economics and have recently been successfully applied to space missions. For example, a market-based approach was used during Cassini spacecraft development from 1993 to 1995 to control the science instruments' demand for resources. Results from the Cassini Resource Exchange (CRE) showed that instrument cost growth was less than 1% and instrument mass growth decreased by 7%. Prior missions usually had mass and cost growth that exceeded well over 100%.

The CRE system was transferred to Southern California's Air Quality Management District (AQMD) and is currently being used to control smog emissions. Market-based systems have also been successfully tested for the Federal Communications Commission (FCC) Spectrum Auction for Personnel Communication Service Licenses, a prototype system developed for manifesting Space Shuttle Secondary Payloads.

For LightSAR, a committee-driven approach was compared to a market-based one. A market-based system uses "rights" and "trades" to resolve conflicts, instead of educated guesses made by a subjective third party. Science Investigators are allocated a "currency" for expressing the relative importance of one request over another. This currency, known as Priority Points, is budgeted to the Investigators, who then assign the points to their data takes to define the "worth" of the request. Investigators are free to express the relative importance of their requests and make trades among themselves to enhance their positions. A further advantage is that this market-based system resides on the Internet and allows Investigators to contribute to the development of a resource timeline remotely, no matter where they are located.

II. TEST BED EXPERIMENTS. In order to test the ability of a market-based system to develop an efficient timeline of data takes, a set of controlled laboratory experiments was conducted. The use of experiments to evaluate comparative allocation systems has been a reliable source of scientific data. The methodology of experimental economics is similar to the use of wind tunnels to test airfoil designs.

The main components of an experiment are: (1) defining what is to be allocated; (2) setting individual incentives; and (3) defining the process by which resources are allocated. For LightSAR, we defined fixed-duration data take requests as the resources to be allocated. There were four data takes per orbit and four orbits per planning period. Two planning periods allowed individuals to carry forward any unused Priority Points from Period One to Period Two.

Investigators from the radar community were asked to set up the experiment by defining the value of each request as a function of its scientific or commercial objectives. Table 1 shows an example of how an Investigator might define the value of the data take requests.

In this example, the Dual Polarimetry Investigator ranked data takes and then assigned them a mission value. Notice that in this example, the Investigator gave data takes for Kuala Lumpur and Indonesia a rank of 2. If this information alone were given to the Integrator, they would assume that each location was equally important and would then assign the data take easiest to incorporate into the time-ordered listing. However, Kuala Lumpur had an assigned value of 45, while Indonesia had only 35. Examining the mission value revealed that the two locations were not equal, as it seemed at first; the Dual Polarimetry Investigator did have a preference.

Defining a mission value for each data take has another advantage over a simple ranking — it provides tradeoff information and establishes the relative worth of each data take. In Table 1, for instance, an Integrator would try to incorporate Vietnam (rank=1), followed by either Kuala Lumpur (rank = 2) or Indonesia (rank = 2). However, using the value column, the Dual Polarimetry Investigation would produce a greater mission value if Kuala Lumpur and Indonesia (45 + 35 = 80 points) could be incorporated into the timeline instead of the number-one-ranked request of Vietnam (60 points). This example shows that a simple ranking does not provide enough information to produce the highest value time-ordered listing.

Table 1: An example of Dual Polarimetry data take requests

Location	Orbit Number	Data Take Number	Rank	Mission Value
Vietnam	1	1	1	60
Kuala Lumpur	1	3	2	45
Indonesia	1	4	2	35
Cambodia	2	3	3	10

For the LightSAR experiment, undergraduates from the California Institute of Technology were used as test Investigators. The students' incentive was financial compensation: they were paid as a function of how well they were able to get their data takes into the time-ordered listing. Students were assigned one of the five roles: Dual Polarimetry, Quad Polarimetry, Interferometry, ScanSAR, Hi-Res Strip or Spotlight. Students then bid for particular data takes that would provide the highest values. A typical student's bid might look as indicated in Table 2.

Table 2: A typical bid showing its status and the number of Priority Points

Status	Location	Orbit Number	Data Take Number	Bid (Priority Points)
New	Vietnam	1	1	25

A bid is simply an expression of the level of importance a particular location has to the student Investigator. The higher the number of Priority Points bid, the greater the value of the request to the student. For the experiment, the bidding proceeded in rounds. Once submitted, successful bids could not be retracted. This rule ensured that bids were monotonic and that the process would converge.

Once bids were received from each student, the round was closed and a solutionwas computed that maximized the sum of Priority Points bid and produced a conflict-free schedule. Only then did the next round begin. Students could see if their data take requests were incorporated into the listing or determine the number of Priority Points needed to "out bid" another user's successful data take request. The students could choose to resubmit their bid with a larger number of points or choose some other data take. Once again, when all bids were received, the round was closed and then solved for the greatest point value. The rounds continued until the value of the time-ordered listing did not increase by 10 percent of the value of the previous round. Rounds lasted approximately 5 minutes apiece, allowing many experiments to be run in a relatively short period of time. The sheer number of resulting iterations allowed students to validate the experiment's design, find flaws in the operations and vary initial conditions.

After student experiments were complete, experiments were performed with the science community. These rounds were much longer, with one round in the morning and one in the afternoon. A Science Investigator could log-on to the LightSAR experiment website, evaluate the time-ordered listing, submit bids and then log-off. The conditions for ending the planning period were the same as for the student experiments. That is, the planning period ended when the value of the time-ordered listing did not increase by 10% of the value of the previous round.

One interesting problem was determining how to end an experiment (i.e., the planning period). If not chosen carefully, a poorly designed ending could produce undesirable results. For example, a specific time can be used for the close of a planning period. However, doing so produces the undesirable incentive for all Investigators to wait until the market is about to close before they submit bids. This practice keeps the bids low and rewards those Investigators who are quick, rather then promoting the highest value requests. It is possible to use a random closing time to overcome this shortcoming, although this approach could adversely effect the outcome if the market closed prematurely. For our experiments, we used the "popcorn" method — when the market is “popping,” bids are coming in and the overall value of the listing is increasing. The market closes when no bids are received over a predetermined period of time.

Another experimental factor that poses a problem is that users do not know *a priori* how much to bid for a given data take request. Since successful bids cannot be retracted, Investigators have an incentive not to overbid and therefore submit the smallest amount needed to out-bid the current request, a situation that could produce many small bids and an excessive number of rounds. A Vickrey-type auction was used to overcome this problem. In a Vickrey auction, the winning bid “pays” the runner-up price. Thus, if Investigator A submits a bid for 45 points and Investigator B submits one for 60 points, Investigator B “wins” the data take request and is debited 45 points from their account.

Vickrey auctions provide incentives for users to be forthright about their bids. If Investigator A tried to underbid by submitting an amount that was lower than what they were willing to spend, Investigator B could submit a bid much higher than Investigator A’s and only have to pay Investigator A’s price. Users therefore have the incentive to make bids for the price they are actually willing to pay, which in turn drives the system to a solution faster and reduces the required number of rounds.

III. EXPERIMENTAL DESIGN. We compared a Serial Draft approach to two market-based approaches, a Simple Market and a Priority Market. A Simple Market allows users submit bids with Priority Points. In a Priority Market, users only specify the request’s priority. This priority is a measure of the request’s importance to the user and has an associated multiplicative factor that is applied to the amount of resources requested to determine a bid price for the particular data take request. Table 3 shows the number of experiments performed for the Serial Draft, Simple Market and Priority Market approaches. Results for the Serial Draft approach were obtained by performing Monte Carlo simulations. As the table shows, each approach was performed for two types of cases.

The first case was called the “simple” case, where all students had identical preferences (i.e., the same data take requests with the same mission values). This case was designed to study the most heavily conflicted situation, where all students desired the same data take time. A second simple case was run where all students had diverse preferences. In this case, all data take requests had different values, such that an optimum solution in which each student was able to obtain their high value data takes was possible. This case was done to see if a market-based system could find the optimum solution.

Table 3: Number of experimental runs performed for each allocation method and associated case

Case	Serial Draft	Simple Market	Priority Market
Simple	Monte Carlo	3	3
Trade-offs	Monte Carlo	3	3
Science Team Simulations			1

The second case was called the “trade-off” case. This case was designed to see if users would accept several lower priority data take requests over their prime request. Trade-off information is nearly impossible to obtain in a Serial Draft approach, because trade-off information can only be obtained through detailed questioning of the Science Investigators by the Sequence Integrators. In addition, only the specific questions asked get answered. Sequence Integrators would be hard-pressed to ask enough questions to understand the full trade space. Once trade-offs are made and the market closes, market-based systems move from bidding to the Aftermarket, a commodities market in which Science Investigators trade any of their resources for any of those owned by another. Aftermarkets are very effective in that both Investigators have to agree to the trade in order for the trade to be completed. This opportunity to barter increases the overall mission value of the timeline. A final case was performed with Science Investigators to obtain their opinions about a market-based system and its relevance to their allocation problem. In essence, the Investigators were solicited to find out if they could use this approach and whether the approach performed more satisfactorily than one using a Serial Draft method.

IV. EXPERIMENTAL RESULTS. Figure 1 shows the experimental results when users have identical data take requests. The abscissa axis has the data from the student subjects and a cumulative result. The ordinate axis shows the percent of the mission value obtained using a market-based system as compared to a Serial Draft approach. Thus, a 100% indicates that the same mission value was obtained with a market-based system as compared to the Serial Draft method. Consequently, a percentage greater than 100% indicates that a market-based system produced a greater mission value for that student. With identical data take requests, a market-based system was able to produce results comparable to a Serial Draft method. In addition, for most students, a Priority Market, where they just had to specify a data take priority, did as well or better than when they had to specify a bid price (i.e., a Simple Market). This results from the fact that a Simple Market is less forgiving. Once a bid was accepted in a Simple Market method, it could not be retracted. A bid with an excessively high number of Priority Points would be accepted and would therefore reduce the student’s authority for making subsequent bids.

In a Priority Market that uses a Vickrey pricing strategy, aggressive bids “paid” the runner-up price. Thus, there was a natural mechanism for preventing excessively priced bids. Only the Priority Points needed to “win” the request were debited from the student’s account, allowing the individual to use his other remaining points for future bids. Figure 2 shows the experimental results when users have diverse data take requests. Here again, a market-based system was able to produce results comparable to a Serial Draft method, and a Priority Market did as well or better than a Simple Market.

In the Simple Case, both with identical and diverse preferences, results reveal that when few conflicts exist, the market-based approaches (i.e., Simple Market and Priority Market) yield results similar to those of the Serial Draft method. That is, market-based approaches were able to find solutions that were comparable to the type of results produced by Sequence Integrators.

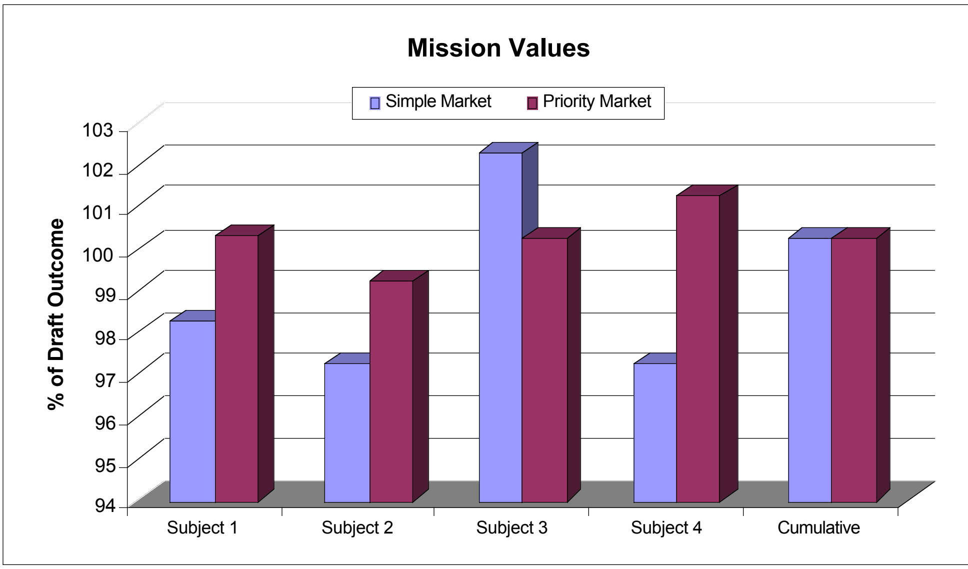


Figure 1. Percent value of market-based approaches compared to a Serial Draft approach, given students with identical data take requests.

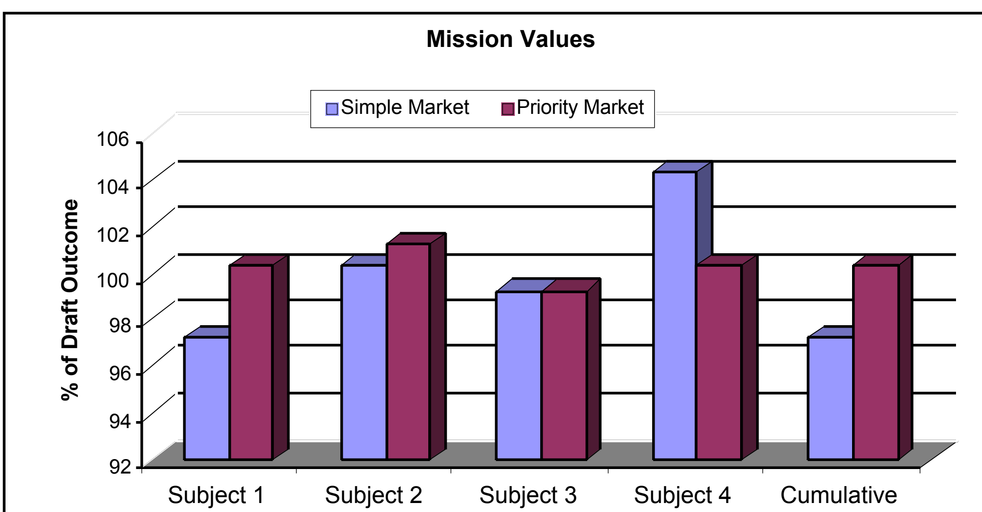


Figure 2. Percent value of market-based approaches compared to a Serial Draft approach, given users with diverse data take requests.

Experimental results for the case with "trade-off" information and "trades" are shown in Figure 3. As can be seen in the figure, Investigators had much to gain by making trades. A significant increase in mission value can clearly be realized by selecting a greater number of lower priority data take requests over a few higher priority requests and/or by carrying forward Priority Points. Thus, when there are trade-

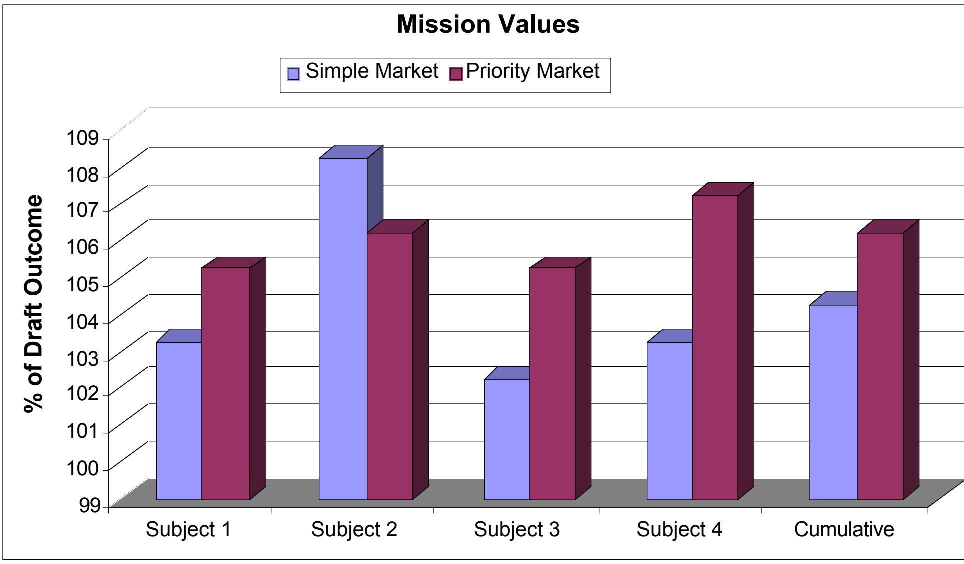


Figure 3. Mission value of market-based approach compared to a Serial Draft approach, given that students could take advantage of trade-offs.

offs in the number and types of data takes (i.e., when significant conflicts exist) a market-based approach produced a greater mission value than the Serial Draft approach.

The results from this case were superior to a Serial Draft approach — a Priority Market produced a 2% greater value than a Simple Market method. In addition, a Priority Market converged to a solution in about half the number of rounds needed in the Simple Market approach. This indicates that a market-based system, using a Priority Market, will arrive at a solution faster than a Serial Draft approach while utilizing fewer individuals to get the same caliber of results.

In addition, using a Vickrey auction Priority Market revealed that specifying a request's priority was more natural to Investigators than specifying a bid "price." The Vickrey pricing strategy made the system more forgiving of bids that may have been too high and motivated individuals to submit bids that honestly reflected their true desire for a particular request. Thus, the Priority Market was easier to use, encouraged the generation of accurate bids, and produced the desired conflict-free, time-ordered listing in half the time of a Simple Market approach.

Experimental results with Science Investigators from the Jet Propulsion Laboratory revealed that there were few operational problems using a Priority Market-based approach. There were, however, a number of concerns that were associated more with the experiment than with weaknesses in the market-based system. These concerns included worries over who determines the initial allocation of points, the questions about whether the experiment was realistic enough (i.e., not enough resources allocated, not enough data takes, etc.), and uncertainties over how long each planning period should be. These issues do not invalidate a market-based system, but accurately reflect the rudimentary capability of the experimental system as compared to one that would actually be used during operations.

Considering that their overall results indicate a market-based system outperforms a simple ranking approach, the LightSAR Project is moving ahead with the development of a prototype web-based planning tool. This tool will have a realistic Science Investigator interface for submitting data take requests, as well as a market-based solver (with a Vickrey pricing strategy) for developing a conflict-free time-ordered listing that can be directly converted into spacecraft commands for operations.

V. PROTOTYPE WEB-BASED LIGHTSAR PLANNING TOOL. The following screens are printouts from an electronic prototype mission planning tool being developed for LightSAR. Though the tool is operational, its rudimentary capability reflects the current immaturity of its development, not its full utility to the Project.

Based on experience from past radar missions, the LightSAR Pre-Project recognized the utility of a graphical interface for the input of data take requests. As such, a Mercator map projection of the Earth was selected for Investigator input. Figure 4 shows the interface with a low resolution map for testing purposes. Notice that the spacecraft's ground tracks are projected onto the map. Only those ground tracks that occurred during the current planning period were displayed.

To use this tool, a Science Investigator interested in a particular data take request would first define a target region on the Earth. This was accomplished by completing the Data Take Request form (see Figure 5). Notice that a given Investigator had to define an imaging mode, experiment name, look angle, and the latitude and longitude of the target region. Only the experiment name and default data take priority were optional fields.

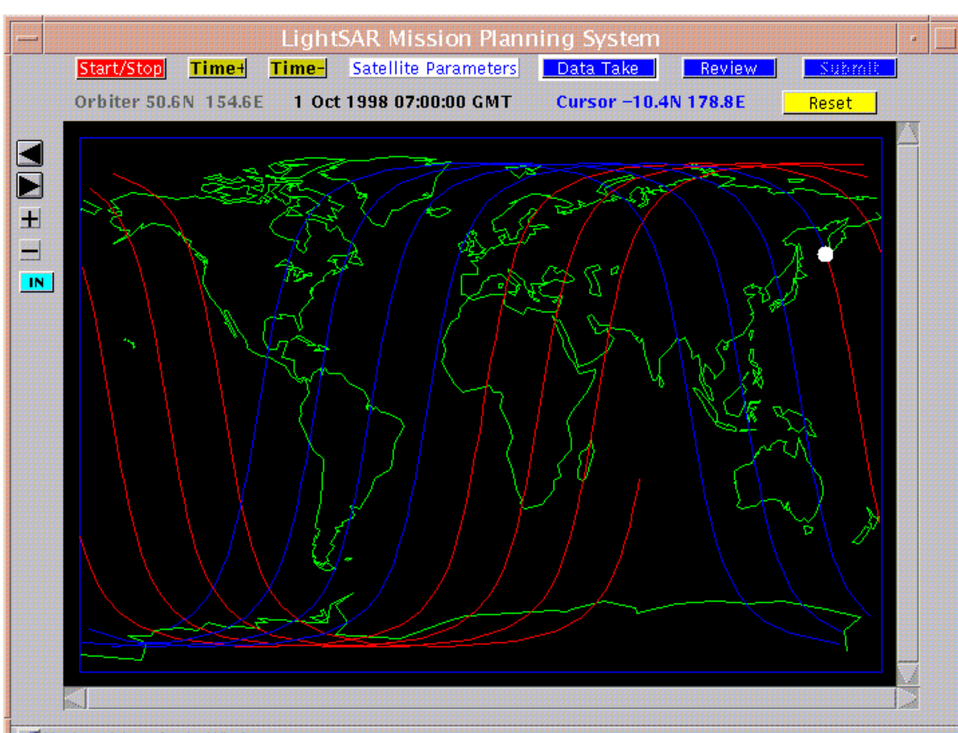


Figure 4. Low-resolution Mercator map of the Earth with projected spacecraft ground tracks.

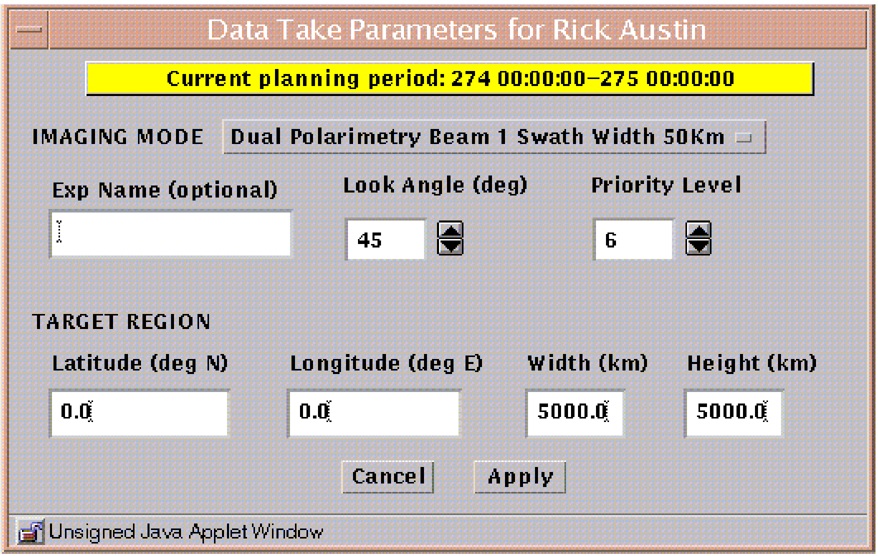


Figure 5. The electronic data take request form.

Once the apply button was selected, the map would show the target region and the associated footprints crossing that particular region (see Figure 6). These footprints were those associated with the ground tracks displayed on the Mercator map.

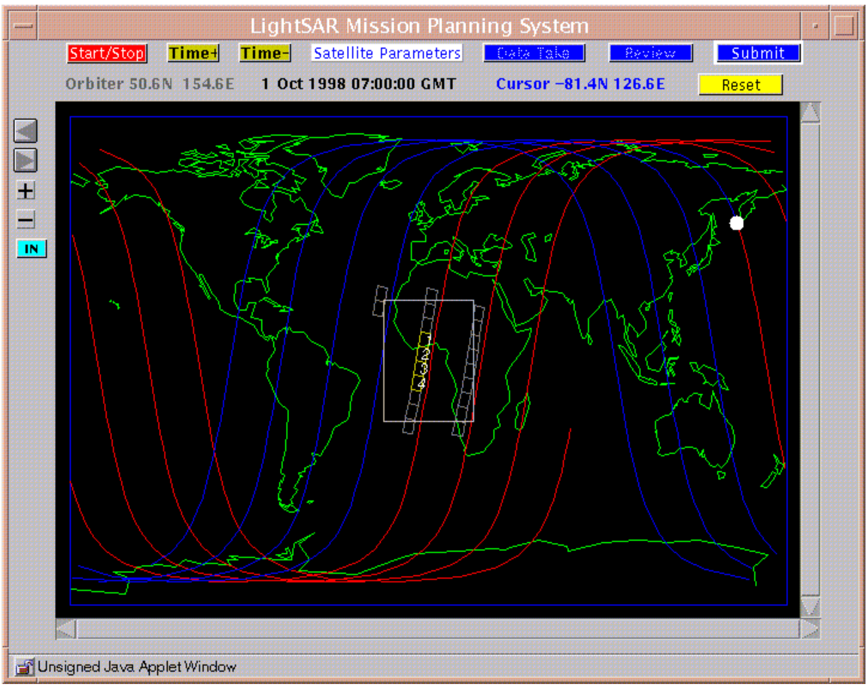


Figure 6. Mercator map of the Earth showing the target region, the associated footprints and the selected numbered footprints.

Next, the Investigator selected which of the footprints in the target region were of interest. These footprints were selected with a mouse and numbered according to the time in which the spacecraft would collect them. The Investigator then clicked on the submit button. The interface returned with the priority selection form (see Figure 7), which showed the Investigator the start and end times for each of his selected footprints and their (defaulted) priorities. The Investigator could than change the priority for any and all of the footprints and use Boolean operators to "and/or" the footprints together. When the Investigator is satisfied with his request, choosing the "submit" button will transmit the request to the Solver.

The Solver returned with a graphical timeline that was conflict-free. (Notice that the timeline is divided into two halves; see Figure 8.) The top half, in light gray, shows the conflict-free timeline. The bottom half, in dark gray, shows unsuccessful bids.

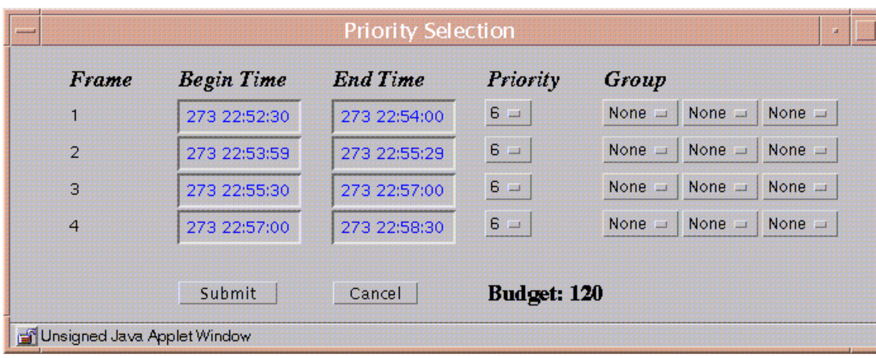


Figure 7. Priority selection form with the data take requests.

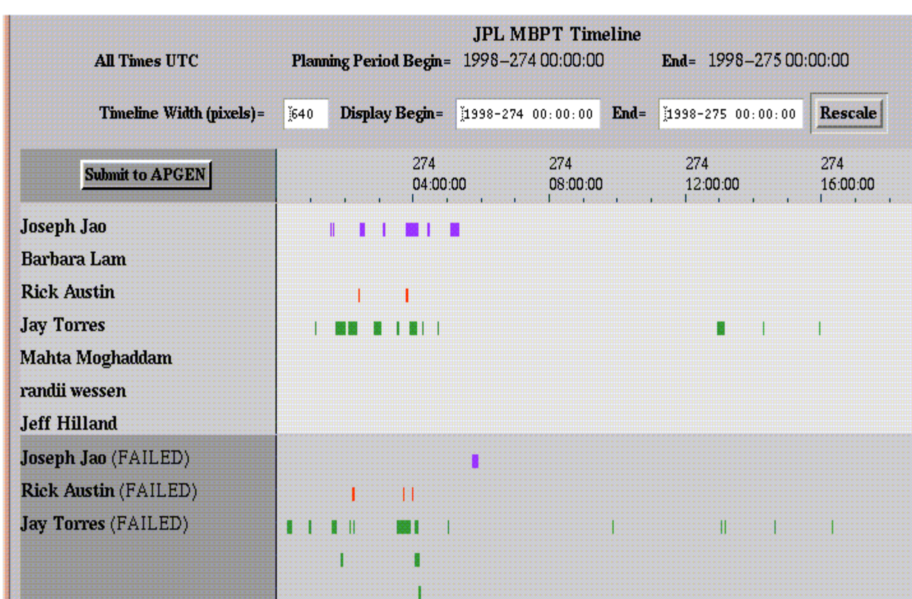


Figure 8. Timeline returned by the Solver of radar data take requests.

From this timeline, Science Investigators can tell whether their requests are successful or not. If unsuccessful, they can either move the time of their request (to avoid the conflict), or increase their bid to try to "win" the time slot.

Once the mission value of the timeline did not increase by 10%, the timeline was considered "done" and ready for the Aftermarket. Pressing the "APGEN" button converts this data to a file that is compatible with the command generation programs used to command the spacecraft. At this point, Science Investigators can trade resources among themselves to increase the value of their data take requests.

Conclusion

Market-based approaches for mission planning outperform the standard Serial Draft approach when there are conflicts between numerous data take requests. Furthermore, market-based approaches: (1) provide clearer priority information; (2) remove conflicts (the timeline is always conflict-free); (3) provide easy access (via the Internet); (4) remove the need for Timeline Integration Meetings; (5) reduce the number of appeals made by Investigators.

The LightSAR experiments show that there are no technical issues associated with the operations of a market-based planning tool. Though there are still some management issues to resolve (e.g., Who allocates the Priority Points? Do Project personnel have the right to veto Investigator trades? How can the graphical interface be made easier to use?), market-based systems have many strengths. They remove the need for Sequence Integrators, are faster (appeals and Integration Meetings are no longer necessary) and can be done remotely from the Investigator's home institution. However, the most important benefit is that that they move the decision-making process back to where the information resides — to the Investigators themselves.

Currently, the market-based planning tool is being evaluated further by the LightSAR Program. It is envisioned that once the Program's industry partner sees the merits of such a system compared with those used on past radar missions, a Web-based, market-based system with a user-friendly graphical interface will be the obvious choice when LightSAR launches early in the next millennium.

Acknowledgment. The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

References

1. TULLOCK, G. (1980), "Efficient Rent Seeking," *Toward a Theory of the Rent-Seeking Society*, edited by James Buchanan, Robert Tollison and Gordon Tullock, College Station: Texas A&M University Press.
2. WESSEN, R. R. and D. PORTER (1998), "Market-Based Approaches for Controlling Space Mission Costs: The Cassini Resource Exchange," *Journal of Reduced Mission Operations Costs*, Vol. 1, No. 1, pp. 9–25.
3. WESSEN, R. R. and D. PORTER (1999), "A Market-Based Approach for Manifesting Shuttle Secondary Payloads," *Journal of Spacecraft and Rockets*, Vol. 36, No. 1, pp. 142–147.
4. VICKREY, W. (1961), "Competitive Sealed Bid Tenders," *Journal of Finance*, Vol. 16, pp. 8–37.

A Modeling Strategy for the NASA Intelligent Synthesis Environment

Paul A. Fishwick

Abstract

We overview the goals of NASA's Intelligent Synthesis Environment (ISE) from the perspective of system modeling. Some of the problems with present day modeling are discussed, followed by a suggested course of action where models as well as their objects are specified in a uniform representation based on the Virtual Reality Modeling Language (VRML). Existing dynamic modeling techniques tend to be 2D in form. The Rube methodology and application provides a 3D modeling framework where model components are objects, and all objects are defined in such a way that they can be easily inserted within World Wide Web documents. This approach suggests the formation of reusable digital objects that contain models.

I. MODELING THE FUTURE. NASA is reinventing itself with respect to new challenges that will culminate in more frequent, less expensive missions. A recent paper by Goldin, Venneri and Noor [10] covers some of the sweeping changes that are to engulf NASA and project it well into the next century. The ISE represents a key technology of the new NASA. How can we begin with a conceptual design for a new spacecraft and take this design through the stages of analysis, testing and fabrication while maintaining the highest level of quality? We enable ourselves to step through sections of the gauntlet with ease if we can generate effective modeling methods. Modeling represents a significant part of ISE since it is with modeling that digital representations of spacecraft are born. ISE is divided into five elements. While the general role of modeling is pervasive in all areas, it is strongest in the ISE elements Rapid Synthesis and Simulation Tools and Collaborative Engineering Environment. It is certainly cheaper to build a virtual spacecraft for Cassini or the Deep Space missions than to construct the actual hardware. And yet, modeling is not without its problems. Modeling can be extraordinarily complex, both in representational schemes and in the iterative procedures required to evolve models over time. My goal is to focus on the modeling aspect of ISE and recommend specific changes in how we design dynamic models that blend seamlessly with the 3D objects being modeled. Models do not have material components; they are ethereal and "live" inside the computer. It is through the practice of modeling that NASA will jumpstart itself into a more efficient future.

NASA Centers are embracing the goals of ISE. Kennedy Space Center [4] has the virtual Shuttle operations to support ground processing. JPL is improving the approach to engineering spacecraft from design to fabrication. The Develop New Products (DNP) initiative has generated significant research in methods for improving engineering design and the processes associated with design. Smith [20] and Wall, et al. [21, 22], define approaches to modifying the existing NASA engineering design practices through model-based means. They point out that much of what exists today in NASA reflects a "document-based" approach to design. A model-based approach is a significant step toward a more manageable process. A related problem is where code is used instead of models. Recognizing the need to surface models will naturally lead to more effective and cost-efficient simulations where the code is automatically compiled, translated from the model's structure. The DNP design cycle is typically divided into the following processes: Mission and System Design (MSD), Design, Build and Test (DBAT), Validate, Integrate, Verify and Operate (VIVO) and Project Leadership and Planning (PLP). Rather than being sequential, these are concurrent and hierarchically related processes with PLP being on the top and proceeding downward to the lower levels of administrative detail as follows: *PLP ⇨ MSD ⇨ DBAT ⇨ VIVO*. The mission is the top-most concern of any NASA process after a project is created. The mission defines what tasks are to be done, and in what order. Sample-based missions involve the collection of material from a comet or a planet's surface. A mapping-based mission would map the surface of a planet or its satellites. Most missions are multi-facetted. For example, Cassini involves flybys and mapping of planets, a moon of Saturn as well as instrumentation for atmospheric experiments for the released Titan probe. The DNP goal is to build cross-cutting (XCUT) models that span all aspects of the mission. If a mission begins with a modeled mission and modeled spacecraft then there will be easier and more effective collaboration among designers, engineers and manufacturing staff. Off-the-shelf commercial software for data flow diagrams and state-based diagrams have been used recently for elaborating modeled spacecraft subsystems.

The ISE goal is "to develop the capability for personnel at dispersed geographic locations to work together in a virtual environment, using computer simulations to model the complete life-cycle of a product/mission before commitments are made to produce physical products"[15]. This is an ambitious goal, but it is on target with the increasing use of modeling and simulation to improve the efficiency with which we design and manufacture components, machines, aircraft carriers, process plants and spacecraft. It also builds upon existing NASA projects (i.e., DNP) that attempt to steer engineering beyond paper and documents to digital representations of objects. In short, we need to use today's cheap computer technology to manufacture virtual equivalents of what we buy and sell.

One of the problems with DNP is that it uses a centralized parameter database, around which programs are situated so that each program reads-from and writes-to the database. This central hub-spoke approach is an improvement over having separate programs each with separate data files and repositories; however, a cleaner approach is to create an object-oriented scene where all data are associated with the relevant objects. The central database approach [9] was also used for the NASA Integrated Programs for AeroSpace Vehicle Design (IPAD) Project [8]. IPAD used a relational database to store structure-based parameters to be used by CAD and Finite Element programs at NASA Langley Research Center. This was a dramatic improvement over separate data files, but it suffered from the fragmentation of connecting data and model to the encapsulating object. With the design and creation of a spacecraft, the scientists and engineers will interact and focus on the physical item — the spacecraft itself. If we can create a process where we build a completely digital spacecraft, then we will maintain this necessary collaboration among all programmers, modelers and engineers. Parameters of a high-gain antenna should be made available to the engineer who touches the antenna; the parameters need to be stored within the objects they define. Moreover, all information about the spacecraft should be so oriented that programs and models are accessible via the digital spacecraft. The spacecraft itself becomes the primary interface for all related models, programs and data. Higher level abstract concepts, such as the mission, can be materialized into objects that remind us of the basic mission elements.

II. PROBLEMS IN MODELING. There are a number of problems that must be addressed once we begin to model. These problems are by no means intrinsic to NASA. They are general problems of the larger modeling community. Even though a large segment of the engineering community acknowledges the importance of modeling, the overall process of modeling is not without its share of defects. Significant changes need to be instrumented if we are to make modeling effective, for if it is not a truly economic enterprise, modeling and simulation will always be seen as choices of last resort, or "to be performed only when time and resources permit." Let's highlight important modeling issues:

- *Modeling Freedom:* Many types of modeling exist, with mathematical models being only one type. We need to reemphasize that models are for humans — not computers. Therefore the models must appeal to the human senses to be effective.
- *Modeling vs. Validation:* The aspects of modeling that we discuss are based on design principles. Even though models are said to be "good" when they validate physical phenomena, all models are flawed in this sense — the model shows us a window of valid behavior of an object and we use it to augment our intellectual and otherwise mathematical methods. The Bohr billiard ball model of the atom is still very useful when used correctly even though we realize that billiard balls are not to be taken literally [11,1]. Validation is separate from modeling but is to be used in conjunction with it. Modeling is what we do to understand and reason about a thing. Validation is taking a model and comparing the model's prediction with experimental results.
- *Code vs. Model:* Too frequently, when one speaks of a model, one is referring to an abstract representation that bears little or no formal relation to the computer code that is supposed to represent the model. While code can be viewed as a model in its own right, more common model forms are based on both equational and highly visual structures. It is essential to have generated code be derived from the model so that all interaction is directly through the model and the code can be forgotten — code represents the cement, whereas the model represents the multi-tiered building created from the cement.
- *Programs vs. Models:* How do computer programs and models relate? The differences between programming, as we generally learn it in universities, and modeling reflect a gradual change in our software and hardware technologies. Computer science and engineering stands out as being separate from other engineering disciplines in the sense that everyone else talks about matter and physics and computer scientists talk of data structures, procedures, relations and objects. Programming has evolved with a heavy bias toward mathematical representation. The problem is that this sort of representation bears little direct connection to physics or to other engineering disciplines. Fortunately, movements are underway in many computer

science areas that suggest alternate, more physical, representational structures [17,2,19].

- *Integration.* NASA is in need of truly integrated virtual, 3D environments where the objects to be modeled, as well as their models, live in the same space. To determine the dynamics of the Cassini probe destined for Titan, one need only touch the 3D probe (attached to the orbiter), activate its behavior field object and then navigate the dynamics that are surfaced in a 3D form. Achieving this means that we have to free the process of dynamic modeling from its two-dimensional home where it has been "imprisoned." Humans better understand and reason with environments that are similar to those found in everyday life. Data that define parameters of spacecraft science instrumentation, for example, cannot live in a database by itself. Parameters are part of objects, and the engineer wants to reason and work with these parameters through the virtual objects that the data represent or modify. It may well be that a very low-level underlying database schema supporting such interaction is still needed, but it is critical to maintain the virtual connections to the data through the physical spacecraft components. This might be seen as an issue of visualization or user interface — and it is. The act of modeling is all about developing and fostering sensory appeal between the human and the modeled object. Thus, it becomes impossible to separate the discipline of human/computer interaction from the task of modeling. They are one and the same. The relational or hierarchical database should disappear from view since it bears no relation to the spacecraft.

III. THE NATURE OF MODELING. One physical object captures some information about another object. If we think about our plastic toys, metal trains and even our sophisticated scale-based engineering models, we see a common thread: to build one object that says something about another — usually larger and more expensive — object. Let's call these objects the source object and the target object. Similar object definitions can be found in the literature of metaphors [12] and semiotics [18]. The source object models the target, and so, modeling represents a relation between objects. Often, the source object is termed the model of the target. We have been discussing scale models identified by their source and target objects having roughly proportional geometries. Scale-based models often suffer from the problem where changing the scale of a thing affects more than just the geometry. It also affects the fundamental laws applied at each scale. For example, the hydrodynamics of the scaled ocean model may be different than for the real ocean. Nevertheless, we can attempt to adjust for the scaling problems and proceed to understand the larger universe through a smaller, more manipulable, version.

Later on in our education, we learned that modeling has other many other forms. The mathematical model represents variables and symbols that describe or model an object. Learning may begin with algebraic equations such as $d = \frac{1}{2}at^2 + v_0t + d_0$ where d, v and a represent distance, velocity and acceleration, and where d_0 and v_0 represent initial conditions (i.e., at time zero) for starting distance and initial velocity. These models are shown to be more elegantly derived from Newton's laws, yielding ordinary differential equations of the form $f = ma$. How do these mathematical, equational models relate to the ones we first learned as children?

To answer this question, let's first consider what is being modeled. The equations capture attributes of an object that is undergoing change in space (i.e., distance), velocity and acceleration. However, none of the geometrical proportions of the target are captured in the source since the structure of the equations is invariant to the physical changes in the target. A ball can change shape during impact with the ground, but the equations do not change their shape. If a ball represents the target, where is the source? The source is the medium in which the equations are presented. This may, at first, seem odd, but it really is no different than the toy train model versus the actual train. The paper, phosphor or blackboard — along with the medium for the drawing, excitation or marking — has to exist if the equations are to exist. In a Platonic sense, we might like to think of the equations as existing in a separate, virtual, non-physical space. While one can argue their virtual existence, this representation-less and non-physical form is impractical. Without a physical representation, the equation cannot be communicated from one human to another. The fundamental purpose of representation and modeling is communication. Verbal representations (differential air pressure) are as physical as those involving printing or the exciting of a phosphor via an electron beam.

Figure 1 displays a painting by the Belgian surrealist artist Magritte, which captures the essence of semiotics and reminds us that source and target objects both must exist. The painting includes a phrase in French “This is not a Pipe”. The object is a painting representing a pipe, or more accurately, it is a piece of paper representing a painting that, in turn, represents a pipe. In the same sense as Magritte’s painting isn’t a pipe, likewise, the equations are (source) objects that we interpret as attributes of other (target) objects. We see an equation and think of the target’s attributes. This leaves us with the wonderful thought that when we model, regardless of the type of model, we use different objects to represent the attributes of other objects. It takes some serious practice to imagine that strange ink impressions on paper might actually represent the position of a ball, train or horse, but that is part of the wonder of modeling and of our ability to perform abstraction — any object can be used as a surrogate for another object’s attributes. In this sense, the more abstract a source object in its relation to the target, the fewer attributes will be found to be in common — a scale model of a train preserves geometry under the right scale transformations, whereas the paper and ink (representing equations) preserves none of this geometry. The equations are said to be more abstract than the scale model.

There is one thing to keep in mind regarding mathematical and 2D image-based models. We use them so frequently because of economic reasons and not because they reflect the best and most natural ways to model. Creating a scale model of the ocean is much easier than using the real ocean. But using a piece of paper or a blackboard is even easier. What if one could create virtual 3D spaces with ease on a portable digital assistant (PDA) device? In the far future, we may even approach the environment of the Holodeck as demonstrated in *Star Trek: The Next Generation*. The Holodeck is a physical space where humans enter fully immersive and interactive 3D simulations. What will modeling be like in such an environment? Will we still draw things on paper or will we gesture to each other while forming 3D worlds that appear before our eyes? The ultimate goal of modeling is not that different from what we did in the sandbox. The difference is that now we can make a virtual sandbox.

Figure 1. Painting by Rene Magritte. Is it a pipe or a *model* of a pipe?

IV. RUBE[1]: BUILDING THE INFRASTRUCTURE. Since 1989 at the University of Florida, we have constructed a number of modeling and simulation packages. We'll begin by describing some early packages and proceed toward our development of the Rube environment. The Web will become a repository for objects as well as documents. Our first package was a set of C programs called *SimPack* [6]. SimPack is a collection of C libraries and programs that allow the student to learn how to effectively simulate discrete event and continuous systems. Discrete event simulation involves irregular leaps through time, where each leap is of a different duration. Discrete event simulation requires scheduling, event list data structures and an ability to acquire resources and to set priorities. Continuous simulation involves stepping through time using equal-sized time intervals and is most often associated with systems based on physical laws. SimPack began as a library for discrete event handling and grew to support continuous modeling (with difference, ordinary and delay-differential equation editors). Fully interactive programs were built upon the core routines and inserted into the SimPack distribution. SimPack is widely used by a number of sites worldwide.

By the early 90's, object-oriented programming was becoming increasingly common in simulation. This suggested that we re-engineer part of SimPack to address the advantages afforded by encapsulation, class hierarchies and re-use. In 1994, we announced OOSIM. OOSIM development started with the event scheduling library in SimPack and expanded upon it to make it more robust using C++.

Both SimPack and OOSIM were found lacking in the user-interface area. Most model types used by scientists and engineers are visual. While we can encode such models in text files, the user doesn't really get a good feel for a model unless it is surfaced in a visible form. In 1997, we began development on a fully visual and interactive multimodeling system, OOPM (Object Oriented Physical Modeler) [5]. Multimodeling [7] is the practice of creating a model at one level of abstraction where each model component can be refined at a level below into a model of a different type than

1. The location of the Rube website is http://www.cise.ufl.edu/~fishwick/rube.

the one at the level above it [14]. For example, the state components of a finite state machine can be refined into differential equations (a different model type). OOPM is based on OOSIM and has a large amount of Tcl/Tk code to support the graphical user interface (GUI). A distributed simulation executive (DSX) has also been constructed for allowing functional block model components to be distributed over the Internet, where each block represents a legacy code responsible for an individual simulation. This system has recently been completed.

During OOPM development, we learned a number of lessons. The first lesson was that even though *multimodeling* had been explained with several formal examples, we lacked an implementation plan and had to carefully work out how the scheduling of "multimodel trees" was to done. The second lesson learned was that GUI development was extremely time consuming. Although everyone wants to use a GUI, one must recognize the significant software engineering effort involved in creating a robust interface. What may appear to be very minor problems from the software engineer's viewpoint can turn out to be critical errors from the standpoint of a human-computer interface. We found out that it is often better to have a primitive text-based interface that is robust than a more complex GUI that has even a very small number of user interface anomalies. Users must develop trust in an application if they are to use it with confidence.

In late 1998 we started designing Rube, named in dedication to Rube Goldberg [16], who produced many fanciful cartoon machines, all of which can be considered *models* of behavior. The procedure for creating models is as follows:

1. The user begins with an object that is to be modeled. For JPL, this can be the Cassini spacecraft with all of its main systems: propulsion, guidance, science instrumentation, power and telecommunication. If the object is part of a larger scenario, this scenario can be defined as the top-most root object.
2. A *scene* and interactions are sketched in a storyboard fashion, as if creating a movie or animation. A scene is where all objects, including those modeling others, are defined within the VRML file. VRML stands for Virtual Reality Modeling Language [3], which represents the standard 3D language for the Web. The Rube model browser is made available so users can "fly though" an object to view its models without necessarily cluttering the scene with all objects. However, having some subset of the total set of models surfaced within a scene is also convenient for aesthetic reasons. The modeler may choose to build several scenes with models surfaced, or choose to view objects only through the model browser that hides all models as fields of VRML object nodes.
3. The shape and structure of all Cassini components are modeled in any modeling package that has an export facility to VRML. Most packages, such as Kinetix 3DStudioMax and Autodesk AutoCAD have this capability. Moreover, packages such as CosmoWorlds and VRCreator can be used to directly create and debug VRML content.
4. VRML PROTO (i.e., prototype) nodes are created for each object and component. This step allows one to create *semantic attachments* so that we can define one object to be a behavioral model of another (using a *behavior* field) or to say that the Titan probe is part of the spacecraft (using a *contains* field), but a sibling of the orbiter. Without prototypes, the VRML file structure lacks semantic relations and one relies on simple grouping nodes, which are not sufficient for clearly defining how objects relate to one another.

5. Models are created for Cassini. While multiple types of models exist, we have focused on dynamic models of components, and the expression of these components in 3D. Even text-based models that must be visualized as mathematical expressions can be expressed using the VRML text node. Models are objects in the scene that are no different structurally from pieces of Cassini — they have shape and structure. The only difference is that when one object is modeling another, one interprets the object's structure in a particular way, using a dynamic model template for guidance.
6. Several dynamic model templates exist. For Newell's Teapot (Section 5), we used three: FBM, FSM, EQN and for Cassini (Section 6), we used one: FSM. These acronyms are defined as follows: FSM = Finite State Machine; FBM = Functional Block Model; EQN = Equation Set. Equations can be algebraic, ordinary differential, or partial differential.
7. The act of creative modeling is to choose a dynamic model template for some behavior for Cassini and then to pick objects that will convey the meaning of the template within the scenario. This part is a highly artistic enterprise since literally any object can be used. In VRML, one instantiates an object as a *model* by defining it: `DEF Parthenon-Complex FSM` {...}. In other words, a collection of Parthenon-type rooms are interconnected in such a way that each Parthenon-Room maps to a state of the FSM. Portals from one room to another become transitions and state-to-state transitions become avatar movements navigating the complex. An avatar is a synthetic human under the control of a program or human user.
8. There are three distinct types of roles played by modelers in Rube. At the lowest level, there is the person creating the *model templates* (FSM, FBM, EQN, PETRI-NET). Each dynamic model template reflects an underlying system-theoretic model [7]. At the mid-level, the person uses an existing model template to create a *metaphor*. A Parthenon-Complex as described before is an example of an architectural metaphor. At the highest level, a person is given a set of metaphors and can choose objects from the Web to create a model. These levels allow modelers to work where they are comfortable. Reusability is created since one focuses on the level of interest.
9. The simulation proceeds by the modeler creating threads of control that pass events from one VRML node to another. This can be done in one of two ways: (1) using VRML Routes, or (2) using exposed fields that are accessed from other nodes. Method 1 is familiar to VRML authors and also has the advantage, in that routes which extend from one model component to an adjacent component (i.e., from one state to another or from one function to another) have a topological counterpart to the way we visualize information and control flow. The route defines the topology and data flow semantics for the simulation. Method 2 is similar to what we find in traditional object-oriented programming languages where information from one object is made available to another through an assignment statement that references outside objects and classes. In Method 1, a thread that begins at the root node proceeds downward through each object that is role-playing the behavior of another. The routing thread activates Java or Javascript Script nodes that are embedded in the structures that act as models or model components for the behaviors.
10. Pre- and post-processing is performed on the VRML file to check it for proper syntax and to aid the modeler. Pre-processing tools include wrappers (that create

a single VRML file from several), decimators (that reduce the polygon count in a VRML file) and VRML parsers. The model browser mentioned earlier is a post-production tool, allowing the user to browse all physical objects to locate objects that model them. In the near future, we will extend the parser used by the browser to help semi-automate the building of script nodes.

Rube treats all models in the same way. For a clarification of this remark, consider the traditional use of the word modeling as used in everyday terms. A model is something that contains attributes of a target object, which it is modeling. Whereas equation and 2D graph-based models could be viewed as being fundamentally different from a common sense model, Rube views them in exactly the same context — everything is an object with physical extent and modeling is a relation among objects. This unification is theoretically pleasing since it captures what it means to model, regardless of model type.

V. EXAMPLE 1: NEWELL'S TEAPOT. In the early days of computer graphics (c. 1974-75), Martin Newell rendered a unique set of Bézier surface spline patches for an ordinary teapot, which currently resides in the Computer Museum in Boston. The teapot was modeled by Jim Blinn and then rendered by Martin Newell and Ed Catmull at the University of Utah in 1974. More recently, Russ Fish produced the image of the teapot in Figure 2, which has the nice property of showing the internal and external teapot shape. While the teapot may seem quaint to us now, it has been used over the years as an icon of sorts, and more importantly as a benchmark for all variety of new techniques in rendering and modeling in computer graphics. The teapot was recently an official emblem of the 25th anniversary of the ACM Special Interest Group on Computer Graphics (SIGGRAPH).

Figure 2. Newell Teapot rendering by Russ Fish, Copyright ©1995, University of Utah.

One of our goals for Rube is to recognize that the teapot can be used to generate another potential benchmark — one that captured the entire teapot, its contents and its models. The default teapot has no behavior and has no contents; it is an elegant piece of geometry but it requires more if we are to construct a fully *digital teapot* that captures a more complete set of knowledge. In its current state, the teapot is analogous to a building façade on a Hollywood film studio backlot; it has the shape but the whole entity is missing. In VRML, using the methodology previously defined, we built TeaWorld in Figure 3. As in Figure 2, the split teapot, we have added extra props so that the teapot can be visualized, along with its behavioral model, in a reasonable contextual setting. The world is rendered in Fig. 3 using a Web browser. *World* is the topmost root of the scene graph. It contains a *Clock*, *Boiling System*, and other objects such as the desk, chairs, floor and walls. The key fields in Figure 4 are VRML nodes of the relevant field so that the *contains* field refers to multiple nodes for its value. This is accomplished using the VRML *MFNode* type. The hierarchical VRML scene graph for Figure 3 is illustrated in Figure 4. The scene contains walls, a desk, chair and a floor for context. On the desk to the left is the teapot, which is filled with water. The knob controlling whether the teapot heating element (not modeled) is on or off is located in front of the teapot. To the right of the teapot, there is a pipeline with three machines, each of which appears in Figure 3 as a semi-transparent cube. Each of these machines reflects the functional behavior of its encapsulating object: *Machine1* for *Knob*, *Machine2* for *Water* and *Machine3* for *Thermometer*. *Thermometer* is a digital one that is positioned in *Machine3*, and is initialized to an arbitrary ambient temperature of 0˚C. Inside *Machine2* we find a more detailed description of the behavior of the water as it changes its temperature as a result of the knob turning. The plant inside Machine2 consists of *Tank1*, *Tank2*, *Tank3* and four pipes that move information from one tank to the next. Inside of each tank we find a blackboard on which is drawn a differential equation that defines the change in water temperature for that particular state. The following modeling relationships are used:

- *Pipeline* is a Functional Block Model (FBM), with 3 functions (i.e., machines).

Figure 3. Office scene with Newell Teapot, dynamic model and props.

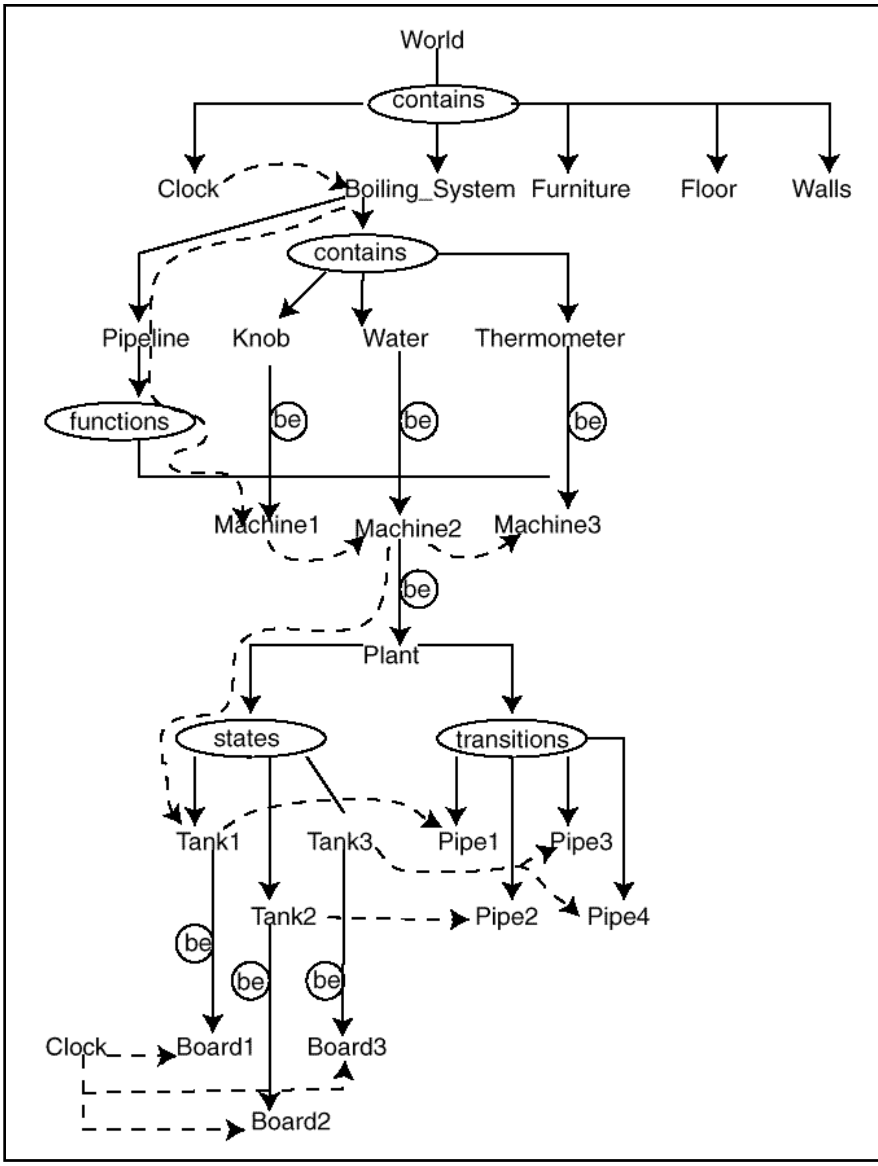


Figure 4. VRML scene graph for the Teapot and its models.

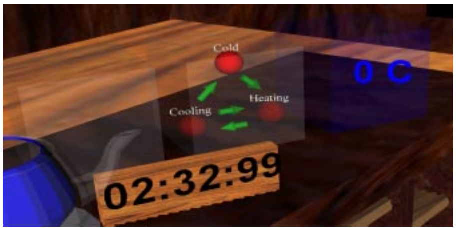
- *Machine* is a function (i.e., semi-transparent cube) within an FBM.
- *Plant* is a Finite State Machine (FSM) inside of *Machine2*.
- *Tank* is a state within an FSM, represented by a red sphere.
- *Pipe* is a transition within an FSM, represented by a green pipe with a conical point denoting direction of control flow.
- *Board* is a differential equation, represented as white text.

The following metaphors are defined in this example. The three cubes represent a sequence of machines that create a pipeline. One could have easily chosen a factory floor sequence of numerically controlled machines from the Web and then used this in TeaWorld to capture the information flow. Inside the second machine, we find a plant, not unlike a petroleum plant with tanks and pipes.

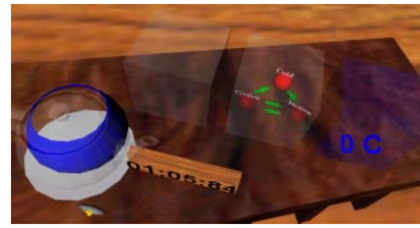
The *Pipeline* and its components represent physical objects that can be acquired from the Web. For our example, we show simple objects but they have been given meaningful, real-world, application-oriented names to enforce the view that one object models another and that we can use the Web for searching and using objects for radically different purposes than their proposed original function. The overriding concern with this exercise is to permit the modeler the freedom to choose *any* object to model *any* behavior. The challenge is to choose a set of objects that provide metaphors that are meaningful to the modeler. In many cases, it is essential that more than one individual understand the metaphorical mappings and so consensus must be reached during the process. Such consensus occurs routinely in science and in modeling when new modeling paradigms evolve. The purpose of Rube is not to dictate one model type over another, but to allow the modelers freedom in creating their own model types. In this sense, Rube can be considered a meta-level modeling methodology.

The simulation of the VRML scene shown in Figure 4 proceeds using the dashed line thread that begins with *Clock*, which has an internal time sensor that controls the VRML time. The thread corresponds closely with the routing structure built for this model. It starts at *Clock* and proceeds downward through all behavioral models. Within each behavioral model, routes exist to match the topology of the model. Therefore, *Machine1* sends information to *Machine2*, which accesses a lower level of abstraction and sends its output to *Machine3*, completing the semantics for the FBM. The FSM level contains routes from each state to its outgoing transitions.

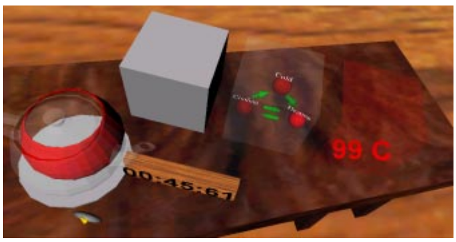
Figure 5(a) shows a close-up view of the pipeline, which represents the dynamics of the water, beginning with the effect of the turning of the knob and ending with the thermometer that reads the water temperature. Figures 5(b)–(d) show the pipeline during simulation when the knob is turned on and off at random times by the user. The default state is the cold state. When the knob is turned to the on position, the system moves into the heating state. When the knob is turned again back to an off position, the system moves into the cooling state and will stay there until the water reaches ambient room temperature at which time the system (through an internal state transition) returns to the cold state. Temperature change is indicated by the color of *Water* and *Machine3*, in addition to the reading on the *Thermometer* inside of *Machine3*. The material properties of *Machine1* change depending on the state of the knob. When turned off, *Machine1* is semi-transparent. When turned on, it turns opaque. Inside *Machine2*, the current state of the water is reflected by the level of intensity of each *Plant*. The current state has an increased intensity, resulting in a bright red sphere. The dynamics of temperature is indicated at two levels. At the highest level of the plant, we have a three-state FSM. Within each state, we have a differential equation. The equation is based on Newton's Law of Cooling and results in a first-order exponential decay and rise that responds to the control input from the knob. The visual display of temperature change confirms this underlying dynamics since the user finds the temperature changing ever more slowly when heating to 100˚C or cooling back to the ambient temperature. Figures 6(a) and (b) show the outside of the heating phase (i.e., red sphere) and the inside of the phase (i.e., blackboard with the first-order differential equation).



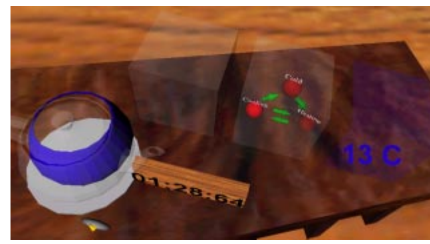
(a) Pipeline close-up



(b) Cold state



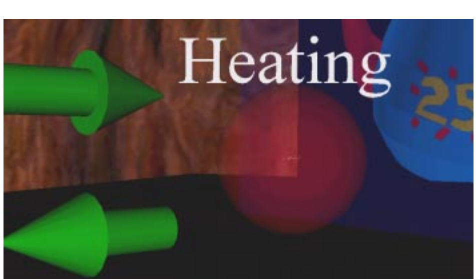
(c) Heating state



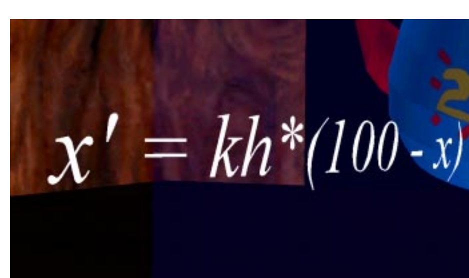
(d) Cooling state

Figure 5. The pipeline behavioral model and the behavioral FSM states defining the phase of the water.

VI. CASSINI. At the time of this writing (June 1999), the Cassini spacecraft has completed one of its Venus flybys. It was launched in October 1997 and plans to make flybys of Venus, Earth and Jupiter on its way to Saturn. Part of the mission is to visit Titan, a moon of Saturn. Figure 7(a) shows a schematic of the Cassini spacecraft while Figure 7(b) shows an illustration of the Huygens probe separation from the spacecraft. The probe descends through Titan's atmosphere and relays science instrument data back to the orbiter. We used the Cassini mission as a basis for a preliminary study on modeling techniques, and we decided to use an FSM dynamic model template to show three phases for the probe: (1) Separation from the spacecraft, (2) Descent and (3) Impact. A scene was created using an architectural metaphor for FSM states. In VRML, the user is located in a room that contains a free-floating model of Titan and Cassini. These models, as well as the model of the room, are visual, computer graphic models meant to act as scaled-down replicas of the actual objects. Scales are non-uniform since Cassini would be much smaller with respect to Titan. The user can freely

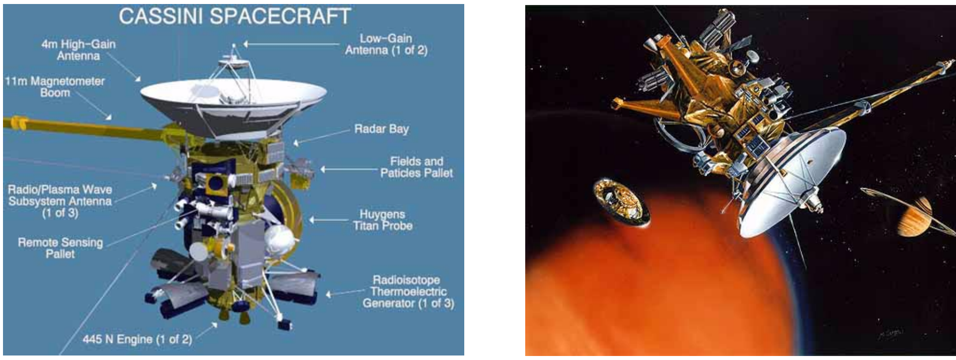


(a) Outside of Heating phase



(b) Inside of Heating phase

Figure 6. Zooming into the Heating phase (*Tank2*).



(a) Spacecraft schematic *(b) Release of Huygens probe*

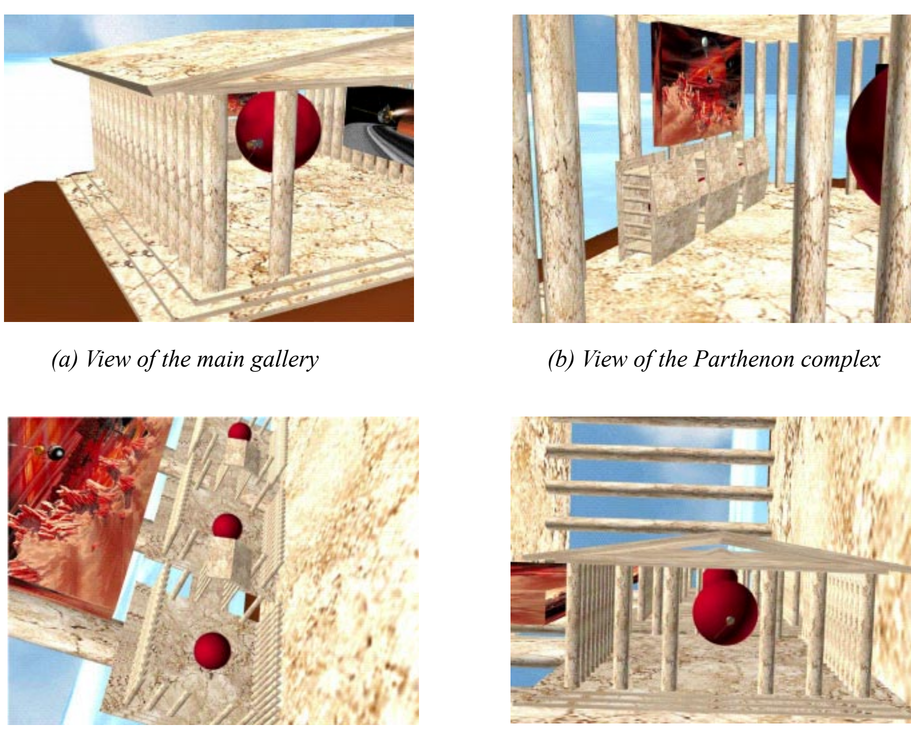
Figure 7. Cassini mission to Saturn and Titan (courtesy JPL).

navigate this environment to view Cassini and Titan. Cassini is shown, with probe attached, making a circular orbit of the moon.

These sorts of visual scale models are common in computer graphics but they represent a small piece of information about Cassini and its mission. Figure 8 displays snapshots of the scene with Figure 8(a) being the Parthenon Room. On three of the four walls of this room, we find color posters relating to the mission. These posters can be selected from within the browser and the user is transported to an appropriate JPL Web page identified by the poster content. Under the poster, in Figure 8(b), is the *Parthenon Complex*, which is an architectural metaphor for an FSM, showing the probe separation in three discrete phases. Figure 8(c) shows three rooms A, B and C). with the following structure: $A \rightarrow B \rightarrow C$. The initial entry room and the three-room environment were created from the Parthenon in Greece. This illustrates an aesthetic aspect of this modeling practice where the modeler is free to choose any type of environment or metaphor. For Cassini, many other types of architectural metaphors come to mind, including the layout of a JPL building or the entire JPL complex (since this represents a common space well known to all JPL employees working on the Cassini project). Even within the confines of the architectural metaphor, there are an infinite number of choices. Within Room A, we may have an avatar that is positioned at the entrance to the room (Fig. 8(d)). There is also a scale model of Titan with Cassini performing the dynamics *associated with the phase* associated with Room A (i.e., probe separation from the spacecraft). Rooms B and C have similar 3D Titan models with dynamics being specified for those phases. The avatar's movement from Room $A \rightarrow B \rightarrow C$ maps directly to the dynamics of probe separation, descent and impact on Titan. The user is able to control the simulation, involving the execution of the FSM, from the main gallery or from inside the complex in Room A. Given this scenario for Cassini, there are some key issues that we should address:

- *Is it a visualization?* The work in Rube provides visualization, but models such as Cassini and Newell's Teapot demonstrate active modeling environments whose existence serves an engineering purpose and not only a post-project visualization purpose for outside visitors. This sort of modeling environment is needed from the very start of a mission — as an integral piece of the puzzle known as model design.

- *Is it economical?* Is this a lot of work just to create an FSM? Why go through the bother of creating the Parthenon, the complex and the avatar? All of these items are reused and so can be easily grabbed from the Web. The concept of reuse is paramount to the Rube approach where the metaphor can be freely chosen and implemented. Without the Web, Rube would not be possible. 3D object placement can be just as economical as 2D object placement, but object repositories are required not only for Cassini and Titan, but also for objects that serve to model the dynamic attributes of other objects (i.e., the Parthenon). Another economical aspect centers on the issue of computational speed for these models. Would creating a simulation in a more typical computer language would be more efficient? The structure of objects and their models within a VRML scene can be translated or compiled into native machine code as easily as source code; the 3D model structure becomes the source code.
- *What is the advantage?* If we consider psychological factors, the 3D metaphor has significant advantages. First, 3D spatially-specific areas serve to improve our memory of the models (i.e., mnemonics). Second, graphical user interfaces (GUIs) have shown that a human's interaction with the computer is dramatically improved when the right metaphors are made available. Rube provides the environment for building metaphors. We leave the ultimate decision to the user group as to which metaphors are effective. A Darwinian-style of evolution will likely determine which metaphors are useful and which are not. Aesthetics plays an important role in metaphor selection as well. If a modeler uses aesthetically appealing models and meta-



(a) View of the main gallery

(b) View of the Parthenon complex

(c) Removing the roof

(d) Side view of the complex

Figure 8. Scene for Cassini and the Huygens Probe dynamics.

phors, the modeler will enjoy the work. It is a misconception to imagine that only the general populous will benefit from fully interactive 3D models. The engineers and scientists need this sort of immersion as well so that they can understand better what they are doing, and so that collaboration is made possible.

- *Is this art or science?* The role of the Fine Arts in science needs strengthening. With fully immersive models, we find that we are in need of workers with hybrid engineering/art backgrounds. It is no longer sufficient to always think 'in the abstract' about modeling. Effective modeling requires meaningful human interaction with 3D objects. So far, scale models have made their way into our engineering practices, but when 'the skin is peeled back' we find highly abstract codes and text. If the internals are to be made comprehensible (by anyone, but most importantly by the engineer), they must be surfaced into 3D using the powerful capabilities of metaphors [13,12]. This doesn't mean that we will not have a low level code-base. Two-dimensional metaphors and code constructs can be mixed within the 3D worlds, just as we find them in our everyday environments (e.g., with the embedding of signs).

At the University of Florida, we have started a *Digital Arts and Sciences* Program with the aim of producing engineers with a more integrated background, thus helping to increase the number of new workers with creative modeling experience[2].

VII. KEY ARCHITECTURAL BENEFITS OF RUBE. The following are novel features of Rube and represent reasons for choosing elements of this architecture:

- *An Integrated Environment:* There is no difference between objects modeling other objects and objects acting in their traditional roles. The modeling and object environments are identical. A pipe can be used in a petro-chemical factory or in a Petri net. Model components are chosen from the vast universe of VRML objects on the Web. Components in models are dynamic as for any object. Models need not be static.
- *Modeling Freedom:* Any 2D or 3D package can be used to create models. There is no need for the Rube team to build a GUI for each model type; the model author can freely choose among drawing and modeling packages.
- *Model Design Flexibility:* There is no predefined modeling method. If a set of objects is to be interpreted as a model, then one adds a small amount of "role playing" information to the objects. Any number of model types can be supported. A side-effect of this flexibility is the provision of natural *multimodeling* support.
- *VRML Encapsulation:* VRML worlds can be stored anywhere over the Web and positioned on an author's Web server through a URL. No new standards have been created outside of existing Web standards, so Rube is built within the framework of VRML, but we can find expressive distributed modeling and simulation capability by "piggy-backing" on the capabilities of the standard. The VRML file that contains prototypes with model fields is a *digital object*, the digital equivalent of the corresponding physical object with all of its attributes. This encapsulation is possible due to the flexible syntax and architecture of VRML (i.e., with key nodes such as PROTO, EXTERNPROTO, Anchor nodes and Sensors being essential for the inclusion of modeling information). The average 3D file standard would leave little room for the definition of models. We propose our modeling methodology as a method for model construction with VRML. In the VMRL community, this has the

2. Refer to: http://www.cise.ufl.edu/fdwi.

potential to alter, for example, how the behavior of objects is modeled. Java and selected behavior scripting languages are currently used, whereas Rube offers the capability for some of this behavior to be modeled and translated into Java using VRML itself to define the behavior.

VIII. REFLECTIONS ON THE ART OF MODELING. It is sometimes difficult to differentiate models used for the creation of pieces of art from those used with scientific purposes in mind. Models used for science are predicated on the notion that the modeling relation is unambiguously specified and made openly available to other scientists. Modeling communities generally form and evolve from a consensual agreement to use certain metaphors. In a very general sense, natural languages have a similar evolution. The purpose of art, on the other hand, is to permit some ambiguity with the hopes of causing the viewer or listener to reflect upon the modeled world. Some of the components in worlds such as Figure 3 could be considered non-essential modeling elements that serve to confuse the scientist. However, these elements may contribute to a more pleasing immersive environment. Should they be removed or should we add additional elements to please the eye of the beholder? In Rube, we have the freedom to go in both directions, and it isn't clear which inclusions or exclusions are appropriate — this choice is entirely up to the modeler or a larger modeling community. One can build an entirely two dimensional world on a blackboard using box and text objects, although this would not be in the spirit of creating immersive worlds that allow perusal of objects and their models.

It may be that a select number of modelers find the TeaWorld room exciting and pleasing. Is this pleasure counterproductive to the scientist or should the scientist be concerned only with the bare essentials necessary for unambiguous representation and communication? Visual models do not represent *syntactic sugar* (a term common in the computer science community). Instead, these models and their metaphors are essential for human understanding and comprehension. If this comprehension is complemented by a feeling of excitement about modeling, this can only be for the better. Taken to the extreme, a purely artistic piece may be one that is so couched in metaphor that the roles played by objects isn't clear. We can, therefore, imagine a kind of continuum from a completely unambiguous representation and one where the roles are not published. Between these two extremes, there is a lot of breathing space. Science can be seen as a form of consensual art where everyone tells each other what an object *means*. Agreement ensues within a community and then there is a mass convergence towards one metaphor in favor of another.

We are not proposing a modification to the VRML standard although we have found that poor authoring support currently exists in VRML editors for PROTO node creation and editing. We are suggesting a different and more general mindset for VMRL — that it be used not only for representing the shape of objects, but all modeling information about objects. VRML should be about the complete digital object representation and not only the representation of geometry with low-level script behaviors to support animation. Fortunately, VRML contains an adequate number of features that make this new mindset possible, even though it may not be practiced on a wide scale. While a VRML file serves as the digital object, a model compiler is also required for the proper interpretation of VRML objects as models.

Summary

There is no unified modeling methodology, nor should there be one. Instead, modelers should be free to use and construct their own worlds that have special meaning to an individual or group. With Rube, we hope to foster that creativity without limiting a user to one or more specific metaphors. Rube has a strong tie to the World Wide Web. The Web has introduced a remarkable transformation in every area of business, industry, science and engineering. It offers a way of sharing and presenting multimedia information to a world-wide set of interactive participants. Therefore any technology tied to the Web's development is likely to change modeling and simulation. The tremendous interest in Java for doing simulation has taken a firm hold within the simulation field. Apart from being a good programming language, its future is intrinsically bound to the coding and interaction within a browser. VRML, and its X3D successor, represent the future of 3D immersive environments on the Web. We feel that by building a modeling environment in VRML and by couching this environment within standard VRML content, that we will create a Trojan horse for simulation modeling that allows modelers to create, share and reuse VRML files.

Our modeling approach takes a substantial departure from existing approaches in that the modeling environment and the object environment are merged seamlessly into a single environment. There isn't a difference between a circle and a house, or a sphere and a Teapot. Furthermore, objects can take on any role, liberating the modeler to choose whatever metaphor that can be agreed upon by a certain community. There is no single syntax or structure for modeling. Modeling is both an art and a science; the realization that all objects can play roles takes us back to childhood. We are building Rube in the hope that by making all objects virtual, we can return to free-form modeling of every kind. Modeling in 3D can be cumbersome and demand considerable patience due to the inherent user-interface problems when working in 3D using a 2D screen interface. A short-term solution to this problem is to develop a model package that is geared specifically to using one or more metaphors, making the insertion of, say, the Parthenon complex rooms a drag-and-drop operation. Currently, a general purpose modeling package must be used to carefully position all objects in their respective locations. A longer term solution can be found in the community of virtual interfaces. A good immersive interface will make 3D object positioning and connections a much easier task than it is today.

There are many unanswered questions concerning the Rube architecture and the effect it may have on the vast community of model authors. For example, many communities have their own internal standards for behavior representation. VHDL (Very High Level Hardware Description Language) is one such community. They have expended vast resources into the use of VHDL. Should they switch to VRML or is there a way that the two standards can relate to one another? We feel that conversion techniques between VRML and the other file-based standards will ameliorate the potentially harsh conditions associated with a migration of standards. Some standards such as HLA (High Level Architecture) do not include a direct provision for model specification since HLA is focused on the execution of distributed simulators and simulations regardless of how they were created and from what models they were translated. In such cases, Rube will provide a complementary technology to aid in the modeling process. Unified Modeling Language (UML) unifies select visual object-oriented formalisms for representing models of software. There is no reason why someone could not build a complete 2D representation using a 2D modeler, such as CorelDraw or AutoCAD, and then construct a grammar to produce the necessary target language code segments needed for UML model execution. Therefore, Rube is a more general procedure for model translation than that provided by most metaphor-fixed visual formalisms. In this sense, the following analogy holds: Rube is to Modeling-Language-X as Yacc is to Computer-Language-Y. Rube is a general purpose model creation facility and Yacc is a compiler-compiler used to create compilers for arbitrary computer language grammars.

We will continue our research by adding to Rube and enhancing it to be robust. In particular, we plan on looking more closely into the problem of taking legacy code and making it available within the VRML model. This is probably best accomplished through TCP/IP and a network approach where the Java/Javascript communicates to the legacy code as a separate entity. We plan on extending the VRML parser, currently used to create the model browser, so that it can parse a 3D scene and generate the Java required for the VRML file to execute its simulation. Presently, the user must create all Script nodes. The model browser will be extended to permit various modes of locating models within objects. A "fly through" mode will take a VRML file, with all object and model prototypes, and place the models physically inside each object that it references. This new generated VRML file is then browsed in the usual fashion. Multiple scenes can be automatically generated.

Acknowledgments. My first thanks go to my students. They are making Rube and the 'virtual sandbox' come alive through their hard work and inventive ideas and solutions. In particular, I would like to thank Kangsun Lee, Robert Cubert, Andrew Reddish, Tu Lam and John Hopkins. I would like to thank the following agencies that have contributed towards our study of modeling and simulation, with a special thanks to the Jet Propulsion Laboratory, where I visited for three weeks during June 1998: (1) Jet Propulsion Laboratory under contract 961427, *An Assessment and Design Recommendation for Object-Oriented Physical System Modeling at JPL* (John Peterson, Stephen Wall and Bill McLaughlin); (2) Rome Laboratory, Griffiss Air Force Base under contract F30602-98-C-0269, *A Web-Based Model Repository for Reusing and Sharing Physical Object Components* (Al Sisti and Steve Farr); and (3) Department of the Interior under grant 14-45-0009-1544-154, *Modeling Approaches and Empirical Studies Supporting ATLSS for the Everglades* (Don DeAngelis and Ronnie Best). We are grateful for their continued financial support.

References

1. IAN G. BARBOUR; *Myths, Models and Paradigms*, Harper and Row Publishers, 1974.
2. GRADY BOOCH; *Object Oriented Design*, Benjamin Cummings, 1991.
3. RIKK CAREY and GAVIN BELL; *The Annotated VRML 2.0 Reference Manual*, Addison-Wesley, 1997.
4. National Research Council; *Advanced Engineering Environments: Achieving the Vision*, Phase I, 1999, http://www.nap.edu/catalog/.
5. ROBERT M. CUBERT and PAUL A. FISHWICK; *MOOSE: An Object-Oriented Multimodeling and Simulation Application Framework*, Simulation, 70(6):379–395, June 1998.
6. PAUL A. FISHWICK; *Simpack: Getting Started with Simulation Programming in C and C++*, 1992 Winter Simulation Conference, pp. 154–162, Arlington, VA, December 1992.
7. PAUL A. FISHWICK; *Simulation Model Design and Execution: Building Digital Worlds*, Prentice Hall, 1995.

8. PAUL A. FISHWICK and CHARLES L. BLACKBURN; *Managing Engineering Data Bases: The Relational Approach*, Computers in Mechanical Engineering (CIME), pp. 8–16, January 1983.
9. PAUL A. FISHWICK, THOMAS R. SUTTER and CHARLES L. BLACKBURN; *Prototype Integrated Design (PRIDE) System: Reference Manual, Volume 2: Schema Definition*, Technical Report 172183, NASA, July 1983; NASA Contractor Report, Contract NAS1-16000.
10. DANIEL S. GOLDIN, SAMUEL L. VENNERI and AHMED K. NOOR; *Beyond Incremental Change*, IEEE Computer, 31(10):31–39, October 1998.
11. MARY HESSE; *Models and Analogy in Science*, University of Notre Dame Press, 1966.
12. GEORGE LAKOFF; *Women, Fire and Dangerous Things: what categories reveal about the mind*, University of Chicago Press, 1987.
13. GEORGE LAKOFF and MARK JOHNSON; *Metaphors We Live By*; University of Chicago Press, 1980.
14. KANGSUN LEE and PAUL A. FISHWICK; *A Methodology for Dynamic Model Abstraction*, SCS Transactions on Simulation, 1996; Submitted August 1996.
15. JOHN B. MALONE, *Intelligent Synthesis Environment: Engineering Design in the 21st Century*, Slide Presentation.
16. PETER C. MARZIO; *Rube Goldberg, His Life and Work*, Harper & Row, New York, 1973.
17. PIERRE-ALAIN MULLER; *Instant UML*, Wrox Press, Ltd., Olton, Birmingham, England, 1997.
18. WINFRIED NOTH; *Handbook of Semiotics*, Indiana University Press, 1990.
19. JAMES RUMBAUGH, MICHAEL BLAHA, WILLIAM PREMERLANI, EDDY FREDERICK and WILLIAM LORENSON; *Object-Oriented Modeling and Design*, Prentice Hall, 1991.
20. David B. Smith and L. Koenig; *Modeling and Project Development*, Fifth International Workshop on Simulation for European Space Programmes – SESP '98, November 1998.
21. S. D. Wall; *Reinventing the Design Process: Teams and Models*, International Astronautical Federation Specialist Symposium on Novel Concepts for Smaller, Faster and Better Space Missions, Redondo Beach, CA, April 1999.
22. S. D. Wall, J. C. Baker, J. A. Krajewski and D. B. Smith; "Implementation of System Requirements Models for Space Missions," Eighth Annual International Symposium of the International Council on Systems Engineering, July 26–30, 1998.

PAUL A. FISHWICK is Professor of Computer and Information Science and Engineering at the University of Florida. He received a B.S. in Mathematics from the Pennsylvania State University, M.S. in Applied Science from the College of William and Mary, and Ph.D. in Computer and Information Science from the University of Pennsylvania in 1986. He also has six years of industrial/government production and research experience working at Newport News Shipbuilding and Dry Dock Co. (doing CAD/CAM parts definition research) and at NASA Langley Research Center (studying engineering data base models for structural engineering). His research interests are in computer simulation, modeling, and animation. Dr. Fishwick is a Fellow of the Society for Computer Simulation (SCS), and a Senior Member of the IEEE Society for Systems, Man and Cybernetics. Dr. Fishwick founded the `comp.simulation` Internet news group (Simulation Digest) in 1987, which now serves over 15,000 subscribers. He has chaired workshops and conferences in the area of computer simulation, and will serve as General Chair of the 2000 Winter Simulation Conference. He was chairman of the IEEE Computer Society technical committee on simulation (TCSIM) for two years (1988–1990) and is on the editorial boards of several journals, including the *ACM Transactions on Modeling and Computer Simulation*, *IEEE Transactions on Systems, Man and Cybernetics*, *The Transactions of the Society for Computer Simulation*, *International Journal of Computer Simulation* and the *Journal of Systems Engineering*. Dr. Fishwick's WWW home page is http://www.cise.ufl.edu/~fishwick and his E-mail address is fishwick@cise.ufl.edu. He has published over 100 refereed articles, authored one textbook (*Simulation Model Design and Execution: Building Digital Worlds*, Prentice Hall, 1995) and co-edited three books.

The First Space Mission Architect: Tsiolkovsky

William I. McLaughlin

Abstract

Konstantin Tsiolkovsky (1857–1935) was the first to develop detailed, realistic designs to support the concept of traveling through space. His three major achievements were: (1) demonstration that a rocket was the unique device that would enable the exploration of space, (2) derivation of the "rocket equation" relating two key design parameters of a rocket to its performance, and (3) use of a systems approach for assembling research elements into a complete design. Author of approximately 500 publications, his core work is presented in the 1903 treatise, "Investigation of World Spaces by Reactive Vehicles." Tsiolkovsky pursued his earlier investigations within tsarist Russia and, after the revolution of 1917, continued working under the Soviet state. His influence was felt within the community of engineers and scientists who designed and launched Sputnik in 1957.

I. INTRODUCTION. The word "architect" is of Greek origin and consists of two parts. The first denotes a leader — as in "matriarch"— while the second refers to the act of building, e.g., "technology." The modern meaning of "architect" is true to these origins; the *Oxford English Dictionary* gives the definition "A master-builder. Specifically, a skilled professor of the art of building, whose business it is to prepare the plans of edifices, and to exercise a general superintendence over their erection."

Konstantin E. Tsiolkovsky (1857-1935), a product of tsarist Russia and far removed from the principal scientific currents of his time, is the unlikely, but undoubted, first in the line of space mission architects. Tsiolkovsky had a clear picture, consistent with the definition of "architect" given above, of what his new profession required: " At first we inevitably have an idea, fantasy, fairy tale, and then come scientific calculations; finally execution crowns the thought." (Sokolsky, 1968, p.84).

Tsiolkovsky's output was large and extended over many decades; his core publication, "Investigation of World Spaces by Reactive Vehicles," concludes that only rockets ("reactive vehicles") would serve for the exploration of space. This treatise was published in 1903, the year of the Wright brothers' first flight at Kitty Hawk. (Prior to his focus on space travel, Tsiolkovsky was interested in aerodynamics and, in 1891, designed and built a wind tunnel.) This publication was more than half a century prior to the actual achievement of space flight, and one might question, on the analogy of the historical cul-de-sac that hid Charles Babbage from the sight of twentieth-century pioneers in computer development, whether Tsiolkovsky's very early work affected the course of astronautics. The answer would be "yes." He became integrated into the community of Soviet scientists and engineers whose efforts culminated in the launch of Sputnik, 100 years and 17 days after Tsiolkovsky's birth. Therefore, even though he was not able to participate directly in a space mission, this lineage would argue that the label "architect" is not misapplied in his case.

II. LIFE. Complete understanding of how an individual attains a high level of achievement would not only require information no biographer could expect to possess, but such insight certainly eludes the biographical subject, too, since our present theories of mind are primitive. Human performance lies only partially within the domain of scientific inquiry. Nevertheless, factors and influences can be examined for their possible effects on the shape of a career. In Tsiolkovsky's case, particularly, such an attempt is needed when one contrasts the magnitude of his accomplishments with his setting in time and place. Three factors influential to his work emerge upon even cursory inspection: 1) psychological consequences of a hearing disability caused by a case of scarlet fever as a youngster, 2) the Russian set of doctrines loosely grouped under the term "Cosmism", and 3) science-fiction romances of Jules Verne.

Hearing loss isolated the young Konstantin from school and encouraged the practice of self tutoring. Tsiolkovsky himself said that his (near total) loss of hearing drove him to excel in order to demonstrate to all that he could rise above his disability.

Nikolai Fedorov (1828/9-1903) is identified as the founder of Cosmism, which, in the words of Hagemeister (1997, pp. 186-7), "…is based on a holistic and anthropocentric view of the universe…[and] its adherents strive to redefine the role of humankind in a universe that lacks a divine plan for salvation…" Fedorov, who was employed in the Moscow library, was an early influence on Tsiolkovsky (Burrows, 1998, p.37) — they were personally acquainted — although there is some question as to the timing of this influence (Hagemeister, 1997, p.197). Tsiolkovsky's practical bent sharpened the fuzzy tenets of Cosmism, as exemplified by three of his quotations:

> "To step out onto the soil of asteroids, to lift with your hand a stone on the Moon, to set up moving stations in ethereal space and establish living rings around the Earth, the Moon, the Sun, to observe Mars from a distance of several tens of versts [1 verst equals 1.067 kilometers], to land on its satellites and even on the surface of Mars — what could be more extravagant!" (Sokolsky, 1968, p.124) This is heady stuff for an engineer in 1911.
>
> "In all likelihood, the better part of humanity will never perish but will move from sun to sun as each one dies out in succession. Many decillion years hence we may be living near a sun which today has not yet even flared up but exists only in the embryo…" (Sokolsky, 1968, p.122)
>
> "The Earth is the cradle of humanity, but we cannot live forever in the cradle."

The influence of Jules Verne (1828–1905), author of *From the Earth to the Moon* (1865) and other novels of science, was acknowledged by Tsiolkovsky as a spur to his imagination. However, he made a point of noting that his use of rocket propulsion was a significant advance over Verne's jarring use of a cannon as a launch device.

Thus, personal adversity, Cosmism, and the inspiration of science fiction seem to have been the principal, visible, driving forces acting throughout the externally uneventful life of Konstantin Edvardovich Tsiolkovsky. This section will conclude with a brief summary of that life; Burrows (1998) and Sokolsky (1968) and the Internet should be consulted for additional material.

He was born on September 17, 1857 in the village of Izhevskoye, about 200 km southeast of Moscow. At the age of 10 Tsiolkovsky contracted the scarlet fever which, as mentioned, resulted in the loss of most of his hearing. Rebounding from the initial depression caused by this misfortune, his interest in invention began to manifest itself through the construction of a series of models: engines, pumps, etc.; he augmented practical faculties with broadly-based studies in mathematics and the physical sciences. Sokolsky (1968, p.307) lists the earliest set of problems which engaged the young student — already his thinking exhibits a bias away from the mundane and toward exploration.

- utilization of the energy of motion of Earth
- shape of a rotating liquid
- high-speed trains
- very long-term balloons
- utilization of exhaust of a steam engine
- use of centrifugal force to sling vehicles into celestial space

An indication of the quality of Tsiolkovsky's thought is his independent, youthful development of aspects of the kinetic theory of gases, a landmark of nineteenth-century physics created decades earlier by Maxwell. His life-of-the-mind culminated, as said above, in the 1903 masterpiece "Investigation of World Spaces by Reactive Vehicles," which was updated by the ever-fertile author in 1911 and 1914.

Tsiolkovsky's biography is, to the first order, the story of his work. He did suffer privation duringWorld War I, but the 1917 Revolution, marking the transition from tsarism to communism, was far kinder to him than to many Russians. In fact, in 1921 he was awarded a pension from the state. The more mystical underpinnings of Tsiolkovskian thought were carefully ignored by the Soviet authorities, and emphasis was placed instead upon the "conquest of nature" facet of his work, which blended well with the party line. He was made a member of what was to become the Soviet Academy of Sciences. The grand old man of the Russian space program continued to work on and publish a series of technical problems through the 1920s and 1930s. He died on September 19, 1935 at age 78.

III. WORK. The technical context for Tsiolkovsky's work is most easily discerned in those areas of his research lying within the broad domain of "celestial mechanics." (We would put his contributions within "trajectory mechanics" today.) The first great synthesis in this field was completed by Pierre-Simon de Laplace (1749–1827) — building on the foundations put in place by Galileo, Kepler, and Newton — in his *Traité de Mécanique Céleste* and was, late in the nineteenth century, superseded by François Felix Tisserand's (1845–1896) masterful 4-volume survey, with the same title as Laplace's treatise. But the courses of natural celestial bodies — stars, planets, comets, etc. — are for the most part well separated from one another, a fact reflected in the content of dynamical treatises. (An exception, "Tisserand's criterion," considers the passage of a comet close to a large, gravitating body such as Jupiter.) Thus, close traverses, crucial for the analysis of launches, landings, flybys, orbiters, were not adequately addressed in the nineteenth-century literature of celestial mechanics. In addition, most calculations were performed under the assumption that the masses of the relevant bodies remained constant. Moreover, the effects of resistive mediums were normally neglected in deference to vacuum.

Tsiolkovsky's premier dynamical achievement was the derivation of the "rocket equation," the starting point in the analysis of the capabilities of propulsion systems. The formulation begins with an application of the conservation of momentum in the case where a small amount dm of mass is expelled with constant relative velocity, v_0

(1) $v_0 dm = (m_0 + m)dv$

where m_0 is the dry mass of the vehicle, m is the time-varying mass of propellant, and dv is the change in velocity of the vehicle. The variables are separable in (1) and, after integrating, some manipulation, and letting m_1 denote the initial mass of propellant,

(2) $v/v_0 = \ln (1 + m_1 + m_0/m_0)$

which is the "rocket equation." The quantity $1 + m_1/m_0 = (m_1 + m_0)/m_0$ is called the "mass ratio." (In (2), v is the velocity of the rocket after the propellant has been consumed.)

The mathematics and physics are so simple as to verge on triviality; the beauty of Tsiolkovsky's conception is that it makes manifest two significant engineering parameters: the exhaust velocity, v_0, and the mass ratio, m_1/m_0. Now it was possible to consider designs for rockets and evaluate competing chemical and mechanical schemes.

Browsing in Tsiolkovsky's papers (Sokolsky, 1968), one can appreciate the wide range of dynamical problems he addressed. They include:

- Escape velocities with respect to various bodies in the Solar System
- Multistaged launch vehicles
- A version of the planetary sphere-of-influence concept
- Effects of atmospheric drag on vehicles
- Weightlessness phenomena inside a vehicle in space
- Shaping of orbits with motor burns
- Complex Earth-Moon trajectories
- Solar pressure
- Observational advantages of polar orbits

As important as his individual studies in dynamics might be, Tsiolkovsky's greatest achievement was his demonstration that only a rocket could serve the needs of space travel. This conclusion is a commonplace to us, but, then, it needed to be established through the employment of rigorous methods. In his 1903 paper, this is precisely what the Russian did, allowing him to state that, for example, the rocket is superior to Verne's fictional cannon. Such thinking moved this ancient device into a new phase of its history, beyond military engagements and entertainment at festivals and into the role of underwriter of cosmic exploration.

The third major achievement of Tsiolkovsky, complementing his work in dynamics and the identification of the primacy of the rocket, was his use of the systems approach. For him, this undoubtedly came to pass because of his unified vision of space travel; it just would not do to have scattered pieces of analysis lying about. Thus, after establishing the rocket as the vehicle of choice, he proceeded to ask himself how that rocket might look and the result was a liquid-fueled system, which has been the workhorse of the space age up to the present. He also examined various propellants with regard to their thermodynamic properties. Beamed energy did not escape his notice, either. He investigated control methods for thrusting and even considered how humans, "cosmonauts," would need to be supported while travelling in space: the possible deleterious effects of weightlessness on the human organism were considered and artificial gravity, through vehicle rotation, prescribed as a cure. Looking forward, from 1926, to the Hubble Space Telescope he said, "This [space environment] is a paradise indeed for astronomers whose chief enemy is the atmosphere, of which there is none here. Astronomers here would make numberless great discoveries with their gigantic telescopes, spectroscopes and photographic equipment" (Sokolsky, 1968, p.162). One needs to read the original papers to see the elements strung together in coherent fashion, but clearly all of this sums to far more than a point mass moving through the void.

We submit, then, that the conjunction of "inventiveness" and "systems engineering" earns the laurel "space mission architect" for this pioneer.

A final note on the corpus of his work, which encompassed some 500 publications, must acknowledge Tsiolkovsky's own ventures into writing science fiction. While he by no means became a force in this literary genre, his publications have not gone unnoticed. *The Encyclopedia of Science Fiction* (Clute & Nicholls, 1993) records the titles of his compositions (along with a list of those which have been translated into English).

IV. AFTER TSIOLKOVSKY. The canonical set of founders of the discipline consists of Tsiolkovsky, the American Robert H. Goddard (1882–1945), and Transylvanian-born Hermann J. Oberth (1894–1989). (This selection, however, constitutes a considerable simplification of the rich historical backgroung, as reference to Ley's classic, *Rockets, Missiles, and Space Travel* (1957), confirms.) Goddard and Oberth, like Tsiolkovsky, were excited by science fiction with its opening of thought to worlds beyond our own.

Goddard's influence upon the development of rocketry was diminished by his reclusive nature, which was undoubtedly fortified by the public ridicule he endured when he speculated that a rocket to the Moon might someday be possible. Unlike Tsiolkovsky, Goddard was an active experimenter with rockets. On March 16, 1926, he flew a liquid-fuel device 184 feet, reaching an altitude of 41 feet. Burrows (1958, p.53) comments "Total flight time was two-and-a-half seconds. Well, the Wrights hadn't done all that much better at Kitty Hawk."

Through his writings and advocacy, Oberth was the best known of the founders and is described as the father of the German rocket and space program. His book, *Die Rakete zu den Planetenraumen* (The Rocket into Planetary Space), published in 1923, achieved international recognition.

Paralleling the accomplishments of these men were the educational and research programs of a new kind of organization: the space society. The cooperative nature of the act of space travel, as it is practiced, makes attending to these roots worthwhile.

The first space society arose in Breslau, Germany in 1927: Verein für Raumschiffahrt, or VfR, i.e., Society for Space Travel. The idea had caught on. The VfR played important roles in the development of German astronautics. Oberth was a member and became president after a time. Willy Ley, also a member (Ley, 1957), gives a comprehensive record of the VfR and its work.

The second major space society to be formed was the American Interplanetary Society, a 1930 creation of G. Edward Pendray and David Lasser. Both Pendray and Lasser wrote influentially on the subject of space travel with Lasser's 1931 *The Conquest of Space*, a pioneering work in the English language. Later called the American Rocket Society, the organization evolved into the present-day American Institute of Aeronautics and Astronautics (AIAA), which is a significant component of the aerospace community, conducting numerous conferences and extensive programs of education.

The third of the great early societies, The British Interplanetary Society, or BIS, was founded in 1933 in Liverpool by P.E. Cleator; its headquarters moved to London soon after. Members have included Arthur C. Clarke and Patrick Moore. After a hiatus during World War II, the BIS was reconstituted in 1945. The form of the postwar society — a leader in space education and advocacy — is in large part due to the intelligence, industry, and fine judgment of Leonard J. Carter, long the Executive Secretary of the organization. The book, *Realities of Space Travel*, compiled by Carter from BIS papers and published in 1957, gives a representation of the state of astronautics at the time when it was poised for the actual leap into space.

References

1. BURROWS, WILLIAM E. (1998), *This New Ocean*, Random House, New York.
2. CLUTE, JOHN and PETER NICHOLLS (eds.) (1993), *The Encyclopedia of Science Fiction*, St. Martin's Press, New York, p.1242.
3. HAGEMEISTER, MICHAEL (1997), "Russian Cosmism in the 1920s and Today" in *The Occult in Russian and Soviet Culture* (B.G. Rosenthal, ed.), Cornell University Press, Ithaca and London, pp. 185–202.
4. LEY, WILLY (1957), *Rockets, Missiles, and Space Travel* (revised and enlarged edition), The Viking Press, New York.
5. SOKOLSKY, V.N. (1968), *K.E. Tsiolkovsky, Selected Works*, Mir Publishers, Moscow.

The Architecture of the CloudSat Mission

Mark J. Rokey, Deborah G. Vane, Ronald J. Boain

Abstract

This paper describes the CloudSat mission, a recent winner of the NASA Earth System Science Pathfinder mission competition. The paper provides a mission overview, describes two fundamental architectural decisions made to meet funding constraints, and concludes with lessons learned from the process.

I. INTRODUCTION. In April 1999, NASA announced the creation of a new Earth mission whose goal is to take unique measurements of clouds and aerosols. This mission, developed under the leadership of Dr. Graeme Stephens of Colorado State University and selected from among 25 other competing Earth science candidates, is called CloudSat.

CloudSat is being developed to investigate how clouds affect climate and to improve weather-prediction models. Clouds exert an enormous influence on our weather and climate. In addition to their key role in the atmospheric hydrological cycle, they dominate the energy budget of the planet through their influence on the Earth's solar and thermal radiation budgets. Clouds cool the Earth by reflecting sunlight back to space and warm the Earth by trapping thermal radiation emitted by the surface and the lower atmosphere. Cloud systems also modulate the pole-to-equator variations in solar insolation, which is the fundamental source of global atmospheric circulation. Because clouds have such a large effect on the Earth's radiation budget, even small changes in their abundance or distribution could alter the climate more significantly than the anticipated changes caused by greenhouse gases, anthropogenic aerosols, or other factors associated with global climatic change.

CloudSat will fill a gap in existing and planned observational capabilities. The technology of current space systems is based on passive sensors, which can only sense the bulk properties of clouds or penetrate the top-most cloud layer. They are unable to accurately measure the altitudes of cloud bases, retrieve ice and liquid water content, or probe the structure of multi-layer clouds. CloudSat will improve validation of weather-prediction models by directly measuring these types of cloud characteristics, which currently are predicted but not confirmed, and will provide the first quantitative, global description of vertical cloud radiative properties.

Perhaps more than any other factor, the architecture of this mission was driven by the cost constraints imposed by NASA Headquarters in this competition. The original estimate for the CloudSat mission was \$185 M, but \$120 M was the maximum amount the competition allowed. These cost constraints led to two significant architectural decisions: (1) the extensive use of non-NASA partners to provide funding for specific portions of the mission and (2) the use of formation flying with another spacecraft to make near-simultaneous measurements with both their payload and ours, thereby removing the need for CloudSat to carry an additional instrument.

II. PARTNERSHIPS . The intention of the NASA Earth System Science Pathfinder (ESSP) Program is to accomplish high-quality, focused, Earth-science measurements by utilizing innovative, streamlined management and implementation approaches. The ESSP Program carries out its science investigations by means of spaceborne observations with capped costs for the entire mission life cycle. These costs include: mission management; spacecraft and instrument definition and development; mission systems integration and test; launch services; on-orbit operations; mission science team support; algorithm development and data processing; data product archiving and distribution; and the publication of results in refereed science journals. [Ref 1.]

ESSP missions must be designed and implemented within tight cost and schedule constraints, and contributions from sources other than NASA are encouraged. In point of fact, accomplishing the CloudSat science objectives within the ESSP funding constraints *required* partnering with other funding agencies, domestic and foreign.

a.) Canadian Space Agency. Within Canada's meteorological and remote-sensing community, there exists a strong interest in the observation of clouds from space. Canadian researchers have been actively involved in many scientific studies relevant to the CloudSat mission, including similar studies done with the European Space Agency. They will contribute to the validation of the CloudSat radar by operating ground-based radars in Canada and conducting appropriate in-situ aircraft observations. Canadian researchers also will participate in the scientific analysis of CloudSat data, emphasizing synergy with other space and ground measurements and the improvement of climate and weather-forecast models.

At the same time, active remote sensing of clouds requires technological developments that are relevant to Canadian industry. In terms of related technology, Canada has the unique capability for 94-GHz high-power transmitter technology and a well-recognized capability in mm-wave RF technology. The Canadian Space Agency will provide the 94-GHz Extended Interaction Klystron (EIK) and the RF front end for the CloudSat Mission.

b.) U.S. Department of Energy. The fundamental goal of the U.S. Department of Energy's (DOE) Atmospheric Radiation Measurement (ARM) Program is to understand and improve cloud and radiation processes in global climate models. This is also the underlying goal of CloudSat, making it a program that complements and contributes to ARM by placing ARM-like observations in a global context. Likewise, the ARM program already supports the research that has lead to a maturing of the techniques and algorithms to be used by the CloudSat mission.

The ARM Program will contribute the core observations for CloudSat algorithm development and validation from its continuous and routine data collection in the Southern Great Plains, the North Slope of Alaska and the tropical West Pacific. These sites are fully instrumented, including ground-based radar and lidar observations and regular aircraft deployments during intensive observational periods.

c.) U.S. Air Force. CloudSat observations have great potential application for defense operations, such as demonstrating the value of cloud radar observations to support operational weather analysis and forecasting. Once demonstrated, cloud radar technologies could be adopted for sustained operational use by the National Polar-

Orbiting Operational Environmental Satellite System (NPOESS). The charter of the Air Force Space Test Program (STP) is to provide spaceflight for a ranked list of DoD experiments, which are reviewed by the DoD Space Experiments Review Board (SERB). (The CloudSat mission was reviewed by the SERB and was included in its FY98 and FY99 ranked lists.) STP and JPL participated in a study to determine the level and type of support for the mission. As a result, the STP program offered to commit funds for CloudSat mission operations. STP will provide the ground data system, including staff and antenna support, through the mission lifetime of 2 years.

d.) Picasso-CENA Mission. A combination of lidar and the 94-GHz CloudSat radar provides significant improvements in the ability to assess cloud radiative forcing and feedback, and is a key element of the CloudSat mission architecture. However, due to the NASA funding constraints of the ESSP-2 program, it is not possible to carry a cloud radar and a lidar on the same spacecraft. The NASA Picasso-CENA mission, planned to launch in the same time frame as CloudSat (March 2003), carries a lidar instrument and flies in loose formation with the Earth Observing System PM satellite (EOS-PM). The desired CloudSat science objectives are accomplished by flying in tight formation with Picasso-CENA.

The Picasso-CENA team participated with the CloudSat team in a preliminary technical analysis of the feasibility of dual launch and formation flying. There are significant cost and technical benefits to a combined Delta launch over two individual Taurus launches. NASA has agreed to co-manifest the two spacecraft. Additionally, this approach enhances the scientific objectives of Picasso-CENA (including joint studies planned between Picasso-CENA and EOS-PM) in addition to those of the CloudSat mission.

III. FORMATION FLYING. Formation flying is a navigational strategy where the separation and relative motions of two spacecraft are controlled to preserve a prespecified geometry. As the funding constraints of ESSP made carrying both a radar and a lidar impossible, the CloudSat team elected to carry a radar and employ formation flying to take advantage of measurements being made by the lidar instrument of the Picasso-CENA mission.

The primary navigational requirement for formation flying with Picasso-CENA is to maintain a specified separation and, as closely as possible, the same groundtracks. For the Picasso-CENA/CloudSat mission scenario, Picasso-CENA will function as the "master" spacecraft and CloudSat the "slave" (or burdened) spacecraft that must react to Picasso-CENA's motion and maneuvers. CloudSat therefore will be responsible for implementing any maneuvers required to maintain this flight configuration.

For optimal performance, it is desirable that the average along-track separation between CloudSat and Picasso-CENA be made as small as practical, so as to be near-simultaneous. Currently, the science team is requiring a mean separation of 442 km (the equivalent of 60 seconds) between measurements.

The second part of the navigation requirement for formation flying has to do with cross-track control, which is equivalent to groundtrack control. It is desirable, with cross-track control, to maximize the amount of overlapping coverage by the radar and lidar footprints. Based on preliminary analyses, CloudSat can control the cross-track variations in groundtracks to a difference of no more than ±450 meters. This translates to overlapping of the radar and lidar footprints approximately 62 percent of the time (assuming a Gaussian distribution for the Picasso-CENA pointing control error).

At the beginning of the mission, when drag effects will be most pronounced, the amount of time between maneuvers to maintain the formation is only 5.4 days (based on conservative assumptions for the atmospheric drag environment). As the mission progresses, the drag environment will become less severe (as the solar maximum subsides) and maneuver frequency will lessen. The propulsive maneuver necessary to restore the formation is approximately 8 cm/sec. Thus, over the life of the mission, no more than 17 m/sec in delta-velocity change will be required for CloudSat to perform formation flying with Picasso-CENA.

IV. MISSION OVERVIEW

a.) Payload. To accomplish the scientific objectives of this mission, CloudSat will carry two instruments, the Cloud Profiling Radar (CPR) and the Profiling A-band Spectrometer/ Visible Imager (PABSI), in addition to sharing observations with the lidar flown on Picasso-CENA.

The CPR is a 94-GHz nadir-looking radar, which measures the power backscattered by clouds as a function of distance from the radar. The CPR will be developed jointly by NASA/JPL and the Canadian Space Agency (CSA), as described previously.

PABSI consists of two instruments packaged together: a high-resolution spectrometer and a 2-channel imager (camera). The objectives of the PABSI instrument are to resolve the O_2 A-band spectrum in the 12950–13130 cm-1 (761.61–772.20 nm) range and to acquire narrow-band images at 747.5 and 761.5 nm (to provide the spatial context for the PABSI spectrometer and the CPR). Both the imager and spectrometer are sensitive to reflected sunlight and thus only generate science data on the dayside of the Earth.

While the CloudSat payload technology is mature, one component of the CPR, the 94-GHz Extended Interaction Klystron (EIK) transmitter, requires space-qualification. The family of EIKs, developed by Communications and Power Industries (CPI) Inc., of Toronto, Canada, has been used extensively in existing ground-based and airborne 94-GHz cloud radars. No fundamental development of new technology is required, but the EIK must be re-packaged for the vibration environment of launch and the thermal environment of space.

b.) Mission Design. The CloudSat mission design is based on the CloudSat spacecraft flying in formation with the Picasso-CENA spacecraft. The precision of this technique enables near-simultaneous, congruent measurements of the same cloud formations as each spacecraft moves along essentially the same groundtrack. This means that many of the mission design options and trades normally available to the Science Team and/or the Systems Engineering discipline are precluded because they have already been decided by Picasso-CENA.

For planning and budgeting purposes, the launch date agreed to by the CloudSat and Picasso-CENA projects is March 2003. This date was chosen because of the availability of the Delta launch complex at VAFB and the projected use of that pad per the launch manifest.

The nominal duration for the Picasso-CENA mission is three years. The nominal duration for the CloudSat mission is 25 months, which allows CloudSat to perform two years of nominal operations starting after one month of on-orbit checkout. During this one-month period, CloudSat will be maneuvered into formation with Picasso-CENA. Also, spacecraft checkout and calibration and instrument calibrations will be performed at that time.

CloudSat and Picasso-CENA has very similar operational orbits. Approximately circular, their orbits have an altitude of 705 km and an inclination of 98.08 deg, very nearly sun-synchronous. The inclination is not exactly sun-synchronous, so the orbit plane will precess slowly with respect to the EOS-PM orbit plane. This slow precession, coupled with the careful selection of the initial ascending node position, gives both CloudSat and Picasso-CENA the opportunity to make coincident radar/lidar measurements with the Moderate-Resolution Imaging Spectroradiometer (MODIS) on EOS-PM. The MODIS field of view, extending 110 deg as seen from the EOS-PM spacecraft, corresponds to a swath width of 2330 km centered on the nadir groundtrack, or the equivalent of a ±10 deg central angle measured on the Earth's surface. Thus, coincident observations from Picasso-CENA and CloudSat will be possible at varying atmospheric look angles from MODIS, assuming the nadir looking Picasso-CENA and CloudSat are over the MODIS measurement swath. The choice of the initial nodal position relative to EOS-PM's node guarantees this condition throughout the mission. The designated orbit inclination causes CloudSat to precess westward at 0.016 deg per day with respect to EOS-PM's sun-synchronous orbit plane.

Table 1: Orbit description and parameters

Description	Parameter
Orbit equatorial altitude (ref.)	705 km
Semi-major axis	7083.14 km
Eccentricity	[a]0.0012
Inclination	98.08 deg
Initial ascending node position	31.06 deg or 14:04 hours wrt to sol meridian
Final ascending node position	20.56 deg or 13:22 hours wrt to sol meridian
Ascending node precession rate	0.9701 deg/day
Argument of perigee position	[a]90 deg
Period	98.88 min
Perigee altitude	717.47 km
Apogee altitude	734.47 km
Min. altitude	705 km
Max altitude	734.47 km
Altitude variation	29.77 km (± 8 km for EOS FF)
Along-track orbit speed	7.052 km/sec
Groundtrack speed	6.755 km/sec
Orbit angular rate	0.0607 deg/sec
Shadow time variations	32.6 – 35.2 min
Beta-angle variations	58.4 – 83.3 deg

V. SPACECRAFT. The spacecraft is a version of the Ball Aerospace RS-2000 commercial line, which has an extensive heritage from ERBS (flying since 1984), RadarSat (flying since 1995) and GeoSat Follow-On (launched early 1998). CloudSat will be the fifth RS2000 spacecraft, preceded by QuikSCAT (Figure 1), ICESat, and the two QuikBird spacecraft.

The CloudSat RS2000 bus will have two significant modifications from its baseline design: a shortened structure and the use of SGLS transponders.

The shortened structure is a result of CloudSat and Picasso-CENA being launched together on a Delta 7420-10 using a Dual Payload Attachment Fairing (DPAF) (Figure 2.) The constraints of the DPAF require shortening of the CloudSat side panel height by 61 cm. However, as the RS2000 was originally designed with an internal payload section and the CloudSat instruments are both external, removing this interior volume can satisfy the reduced height constraint without a major spacecraft redesign.

Using the U.S. Air Force for mission operations requires changing to SGLS-compatible transponders. The CloudSat RF uplink/downlink uses redundant SGLS transponders for data transmission and reception. The uplink command data rate of 2000 bps has a worst-case link margin of 13 dB. Science data is downlinked on a carrier signal at a rate of 5 Mbps, with a worst-case margin of 6 dB. A nadir-pointing patch antenna is used for science data transmission. Stored engineering data is downlinked at a rate of 256 kbps on a second carrier signal, with a margin of 6 dB. Real-time data is downlinked on a subcarrier with the stored engineering data at 16 kbps with a worst-case margin of 5.4 dB. A multiple-patch antenna configuration provides spherical coverage for telemetry, command transmission and reception.A summary of the characteristics of the CloudSat spacecraft is given in Table 2.

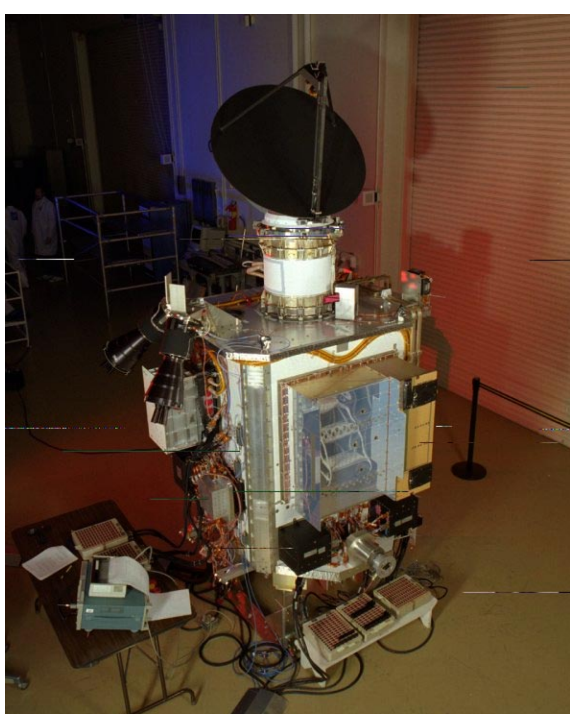


Figure 1. The RS2000 bus, shown in its QuikSCAT configuration.

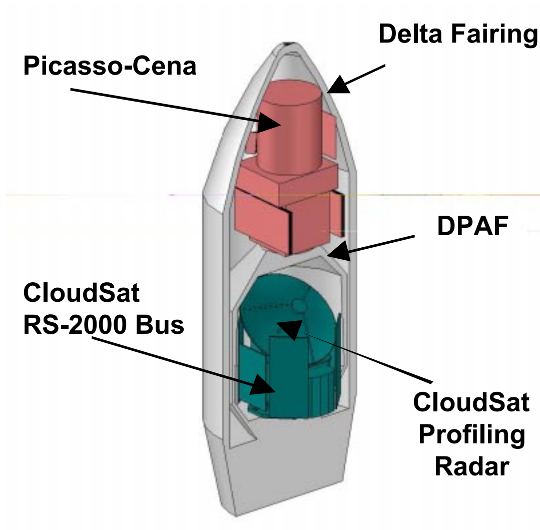


Figure 2. CloudSat/Picasso-CENA dual launch configuration.

Table 2: Spacecraft characteristics

Parameter	**Characteristic**
Design Life	>5 years
Launch Vehicle	Delta 7420-10
Approximate Size	1.9 × 1.9 × 2 m
Mass (wet)	677 kg
Redundancy Approach	Fully Redundant Bus. Ps = >0.95
Control System	3-axis stabilized, zero net momentum, stellar-inertial
Navigation	GPS
Available Power	1375 W EOL
Solar Array Size, Type	6.9 sq.-m, single-axis articulation and s/c yaw maneuvers, dual-junction Ga/As cells
Onboard Data Storage	32 Gbits
Comm. Approach	SGLS, 5 Mbps downlink, 2 kbps uplink
Thermal Control	Primarily passive with some survival heater control

The payload configuration maximizes the size of the CPR antenna while still satisfying the DPAF envelope. The CPR is mounted to an upper deck using semi-kinematic mounts that allows it to be co-aligned with the PABSI. All other spacecraft components and the PABSI are mounted to the exterior surfaces of the shear panels. An exploded view of the spacecraft showing the various components is shown in Figure 3.

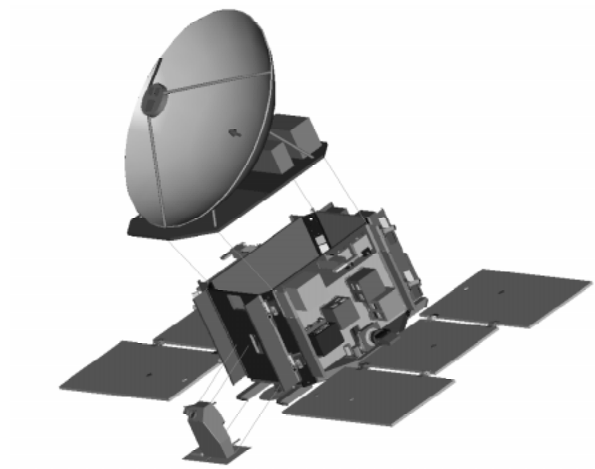


Figure 3. The RS2000 Bus (center), the CPR (above), and the PABSI (below).

VI. GROUND SYSTEM AND MISSION OPERATIONS. The CloudSat ground system uses the existing facilities and personnel of the Project's university and military partners. The USAF Research Development Test & Evaluation Support Complex (RSC) facility at Kirtland Air Force Base will provide flight operations, including mission planning, command generation, telemetry monitoring, spacecraft engineering, level-0 data processing. The Air Force Satellite Control Network (AFSCN) will be used by the RSC for all CloudSat ground antenna support.

Colorado State University's CIRA (Cooperative Institute for Research in the Atmosphere) facility will provide science data processing, which includes levels 1-N data processing, distribution, and data archiving.

The Jet Propulsion Laboratory will provide mission management and the science team interface during operations. A block diagram of the GDS showing the end-to-end concept for operations data flow is shown in Figure 4.

CloudSat's data acquisition strategy is simple and does not vary during the mission. The CPR collects data continuously, while PABSI collects data when the ground beneath it is illuminated by the sun. CloudSat does not need frequent ground contacts from a control point-of-view, as it has no short data latency or adaptive commanding requirements.

The data return strategy was designed to be compatible with a 5-Mbps S-band downlink, so that any of the eight existing AFSCN antenna sites (Figure 5) could be used to support CloudSat in their present configuration. In a typical day, CloudSat would be in view of the AFSCN for more than 4.5 hours, of which only 28 minutes are needed to return an entire day's worth of data — a mere 10% usage requirement. CloudSat will collect 6.7 Gbits of data per day, but carries 32 Gbits of onboard storage, enough for outages lasting several days without losing science data.

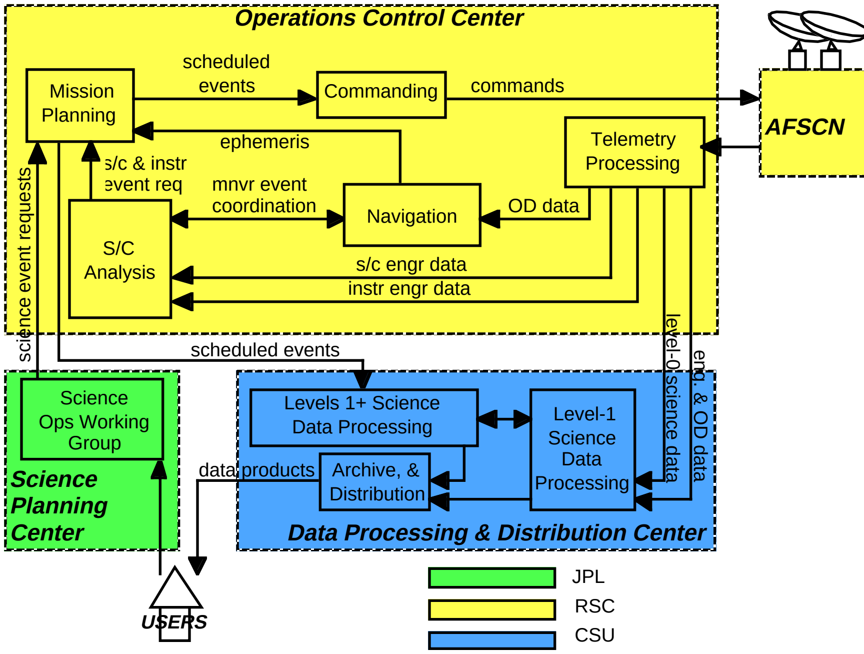


Figure 4. The CloudSat ground system.

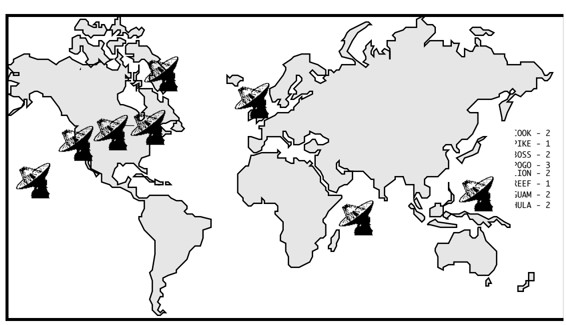


Figure 5. CloudSat is compatible with all the AFSCN global assets, providing maximum flexibility for tracking support.

The operational scenario begins at JPL, where a conflict-free set of high-level science plans will be produced. At regular intervals, these plans will be delivered to the RSC, where they will be used to develop command loads. Real-time operations can be performed 24 hours a day, 7 days a week. The mission control team will consist of mission planners, orbit analysts and contact specialists. Mission planning is performed in mission-dedicated areas, complete with mission-specific documentation, computers and software. The RSC will run four contacts per day using the AFSCN. Once the science data is returned and Level 0 processed by RSC personnel, it will be shipped via tape to CIRA for data processing.

CIRA will process and archive data to produce CloudSat mission science data products for delivery to the NASA DAAC. In addition, CIRA will collect and archive supportive data such as geostationary satellite imagery, synoptic surface observations, upper air observations, and other data required by the Science Team. Because these data are normally available at CIRA, this is provided at no extra cost to CloudSat. The supportive data will be used for input to Science Team applications for the quality control of the CloudSat Level 0 data and subsequent physical interpretation of the CloudSat science data products. The ancillary data will be archived by CIRA, along with CloudSat Level 0 data, using the EOS-DIS HDF format for transfer to the appropriate NASA DAAC.

VII. LESSONS LEARNED

a.) Form Partnerships Early. The development of a partnership requires resources and a great deal of time to accomplish. Multiple interactions with each potential agency are required. Based on experience with CloudSat, our recommendation is to begin this process at least one year ahead of the anticipated proposal date. Some agencies have internal review cycles on an annual basis, and these reviews may be needed for approval of commitments and allocation of resources.

b.) Build a Broad Support Base. The concept for the CloudSat mission grew from the expressed needs of the international scientific community, including weather forecasting centers and the climate research community. The World Climate Research Programme's (WCRP) Global Energy and Water Cycle Experiment (GEWEX) program hosted the first international workshops on cloud radar and communicated to the world's space agencies of the need for spaceborne cloud radar measurements. In a process which lasted over six years, a broad base of support and awareness grew.

Thus, before the CloudSat proposal was submitted to NASA, NASA was aware of the need for such a mission and anticipated the proposal submission. By the time that CloudSat was submitted as an ESSP proposal, there was no doubt that this was a high-priority scientific mission. Letters of endorsement for CloudSat came from the military sector, the weather-prediction centers and the climate research community. Even so, the ESSP competition was intense, with several high-value science missions on the table. Nonetheless, CloudSat was known and strongly supported, a factor which contributed to the outstanding scientific score during the ESSP evaluation process.

Generating this type of support can take years to accomplish. A broad base of support also increases the chances of success in partnering with other agencies.

c.) Buy the More Capable Spacecraft. . As part of the ESSP proposal process, the CloudSat team put out a request for proposals (RPF) that was answered by five aerospace companies who submitted production spacecraft designs. The CloudSat team listed both minimum spacecraft capability requirements and a maximum allowable cost in the RFP. Of the five vendors, two clearly met the guidelines. At this point, the CloudSat team was faced with a very difficult architectural decision. One vendor elected to provide margin in the cost (e.g., a lower price than the cap we set) while meeting the spacecraft requirements. The second vendor elected just the opposite: provide margin in the spacecraft design (e.g., extra capabilities) while meeting the maximum allowable cost.

The cost-constrained nature of this process made the price savings offered by the first vendor quite attractive. However, after careful consideration, the CloudSat team elected to spend the whole of its budgeted amount for the more capable spacecraft. As a general rule, it is our opinion that a project will benefit in the long run by adding margin in their spacecraft design. In recent months, as design details beyond the scope of a proposal effort came into focus, this decision has been proven to be the correct one for CloudSat.

Conclusions

CloudSat was selected in large part because of creative work done to control mission costs. The original, projected cost of this mission was $185 M. The architectural decisions described in this paper reduced NASA's CloudSat cost to $111 M, well below the $120 M cost cap. A summary of these savings is given in Table 3.

The CloudSat mission builds on considerable design heritage and design maturity. This is necessary given the quick schedule to launch, a guideline of the ESSP program, and the need to stay within the ESSP cost cap.

Table 3: Summary of cost savings, resulting from architectural decisions

Decision	Cost reduction to NASA
Use of formation flying to provide LIDAR measurements.	$30 M
Net launch vehicle savings, going from cost of one Taurus launch vehicle a shared launch with a Delta 7420-10 with DPAF.	$20 M
Partner Contributions (USAF contribution of operations, DOE contribution of validation, CSA contribution of radar components).	$24 M
TOTAL	$74 M

Acknowledgements. The research described in this paper was performed by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. The authors would acknowledge the contributions of Randy Coffey of Ball Aerospace, Don Reinke and Ken Eis of Colorado State University, and Sam Sims of the Aerospace Corporation. Our special thanks to Dr. Fernand Rheault (Canadian Space Agency), Lt.Col. Gary Hendel (USAF) and Lt. Dale White (USAF), without whose support we could not have satisfied our cost constraints.

References

1. NASA Earth System Science Pathfinder Announcement of Opportunity – 2, AO-98-OES-01, April 1998.

Satellite Relay Alternatives In a Hyperspectral-Imaging Architecture

Matthew F. Marshall, David A. Bearden, Robert C. Kellogg, Thomas J. Lang

(Extracted from a paper originally presented at the 5th Australian Space Development Conference, Sydney, Australia, July 1998.)

Abstract

High-bandwidth remote sensing systems such as hyperspectral imagers are constrained in the amount of data that can be returned. This is typically due to limited onboard storage or downlink capability (data rate and time) to the ground station during overflight. The study presented in this paper examines the use of low earth orbit (LEO) broadband communication satellites to relay mission data to the ground. A broadband relay system offers the potential of nearly unlimited real-time return of data. However, to minimize the complexity of the communication interface to the satellite relay, the orbit of the imaging satellite or constellation may be constrained. Furthermore, there may be potential outages due to connectivity gaps with the broadband relay system. A comparison with respect to life cycle cost between a traditional store-and-forward approach and a broadband satellite relay approach is provided. The analysis makes use of information from the Teledesic and Celestri systems as they existed in 1998 as representative examples of broadband systems. While Celestri no longer exists as a separate system and the Teledesic design has changed, the conclusions still hold with respect to the issues and potential benefit of a LEO broadband relay to satellite mission architectures.

I. CONCEPT OVERVIEW. The concept of operations and architectural alternatives for this study are shown in Figure 1. The Hyperspectral Imaging (HSI) satellite images approximately 50 areas of interest within the sunlit portion of an orbit, according to previously delivered tasking orders. The HSI satellite is designed for a specific spatial resolution (5 or 10 m) with 64 bands in the spectral range of 0.4 to 2.5 μm.

The HSI satellite will be part of a constellation designed to provide access to every point on the earth at least once per day and may be constrained by the use of a commercial broadband relay system (i.e., need to match nodal regressions with commercial relay constellation in order to simplify the communication interface). Once the raw HSI data has been collected, it will be distributed to a central archive for processing/analysis and release to customers and/or directly addressed to end users using the commercial relay system. Distribution to this central archive can take one of several routes depending on the method chosen for getting the data off the satellite. This study includes all costs for the space and ground segments up to and including the processing of raw imagery data in the central archive.

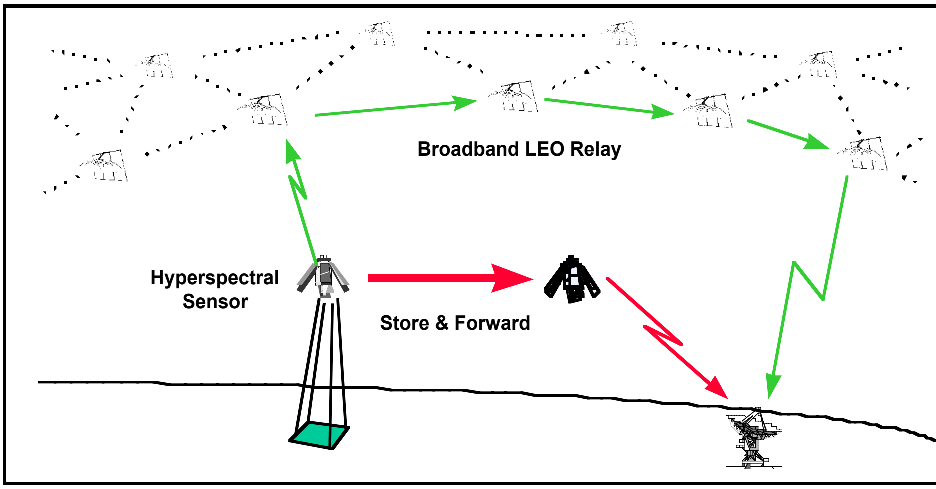


Figure 1. Concept of operations.

II. CASES. There are two main options for the data relay function from the satellite to the ground. These are (1) An RF Uplink to a satellite relay, and (2) Downlink to an existing ground station (store and forward). Two payload ground resolutions (5 m or 10 m) were varied among these options, resulting in a total of four cases, as shown in Table 1. The guidelines for the architecture study are shown in Table 2.

The store and forward cases form the baseline approach for this study, to which the Satellite Relay cases are compared. Thus, there is a corresponding store-and-forward case with equivalent HSI Payload for each of the Satellite Relay cases in order to afford a one-to-one comparison. Since there was no need to match up with a satellite relay, the constellation for the store-and-forward cases were optimized to meet the requirement of revisiting any point at least once per day in daylight. Additionally, high latitude ground stations in Alaska and Norway were assumed for downloading all mission data from each spacecraft.

Table 1: Study trade space

Case	Relay System	HSI Resolution (m)	Data Rate (Mbps)	HSI Altitude (km)
RF Uplink to Satellite Relay				
1	Teledesic	5	622	500
2	Celestri	10	155	500
Downlink to Existing Ground Station				
3	Store-and-forward	5	622	500
4	Store-and-forward	10	155	500

Table 2: Study guidelines

Mission	
Mission lifetime	15 years
Ground lifetime	17 years
Launch date	2005
Technology freeze date	2002
Launch vehicle	Not specified
Constellation	
Coverage	Global coverage over ±70 deg latitude (goal)
Mission orbit	Matched to broadband LEO constellations for Teledesic and Celestri relay cases; optimized for store-and-forward cases
Revisit requirement	Access to any point at least once per day in daylight
Constellation size	Sized to meet orbit and revisit constraints
Payload	
Description	Hyperspectral imager
Resolution	5 m or 10 m, depending on required data rate
Data rate	Separate designs for 155 and 622 Mbps max data rates
Performance	Comparable signal-to-noise ratio (SNR) performance among designs
On-board storage	Up to 50 areas of interest for store-and-forward cases
Costs	
Fiscal year	1998
Cost Target	Not specified (no upgrades during mission life; no new nonrecurring costs)

RF Link to Satellite Relay. The purpose of these cases was to assess the effectiveness of utilizing a LEO broadband satellite relay. Two representative commercial broadband relays based on Teledesic and Celestri were considered. Table 3 presents a summary of the Teledesic and Celestri systems used in this analysis (subsequent to the date the research presented in this paper was conducted, Celestri was canceled and the design of Teledesic was changed). It must be noted that the purpose of this study was not to assess the feasibility of actually making a connection to Celestri or Teledesic, but to examine the utility and benefit of making a connection to a LEO broadband relay relative to a traditional store-and-forward approach. As envisioned, neither Teledesic nor Celestri is designed to allow a satellite-to-satellite link. It is not part of their business plan. Modifications to hardware and/or software would need to be made in order to effect such a satellite connection.

Table 3: Relay satellite characteristics

	Celestri	Teledesic
Number of satellites	63	288
Orbits	7 planes of 9 sats/plane	12 planes of 24 sats/plane
Inclination	48°	88°
Altitude	1400 km	1200–1400 km
User links	Up to 155 Mbps	Up to 622 Mbps

In order to simplify the communications interface, a constraint was imposed such that each of the HSI satellite orbits would remain in-plane to a single plane of the Celestri or Teledesic constellations. This requires that the HSI orbits have the same nodal regression rates as the Celestri or Teledesic orbits. At an HSI altitude of 500 km, this requirement dictates an HSI inclination of about 88 degrees to match Teledesic and about 64 degrees to match Celestri. Communications between the HSI satellites and relay satellites were constrained to these corresponding planes. Out-of-plane communications were not allowed. This was to minimize the variation in relative velocities and hence, Doppler shift, between the sensor and relay satellites. Furthermore, it was assumed that a satellite relay connection could not be made unless the HSI satellite was within the maximum off-axis angle constraint of the broadband relay system. This is more of a conservative operational constraint than a physical constraint.

For the uplink from the sensor satellite to the relay satellite to be closed, the sensor satellite must be within the earth coverage cone of the relay satellite. Figure 2 shows the worst case off-axis (off-nadir) angle at which the Celestri or Teledesic satellite views the HSI satellite, for a range of HSI satellite altitudes. Continuous uplink of data from a sensor to the Celestri constellation would require Celestri to link to the sensor satellite as much as 62 degrees off nadir (at 500 km altitude). However, the Celestri satellites are generally quoted to require a minimum ground elevation angle of about 16 degrees or an off-nadir angle of 52 degrees (Motorola, 1997). Therefore, the question arises as to whether Celestri can link to the HSI satellite as far as 62 degrees off nadir.

If Celestri is truly limited to an off-nadir angle of 52 degrees, then the link from the HSI to the Celestri plane to which it is tied will experience an outage approximately 68% of the time. Because the HSI satellites are in different planes from the Celestri satellites, the handover problem is worse at the higher latitudes than near the equator, because the separation between the HSI satellite planes and the corresponding Celestri plane is greater at the higher latitudes than near the equator. Although there are more Celestri satellites in view of a given HSI satellite at high latitudes, recall that the architecture was conservatively constrained such that a single plane of HSI satellites communicates to only a single plane of nine Celestri satellites, and that the nodal regressions of these planes are equal.

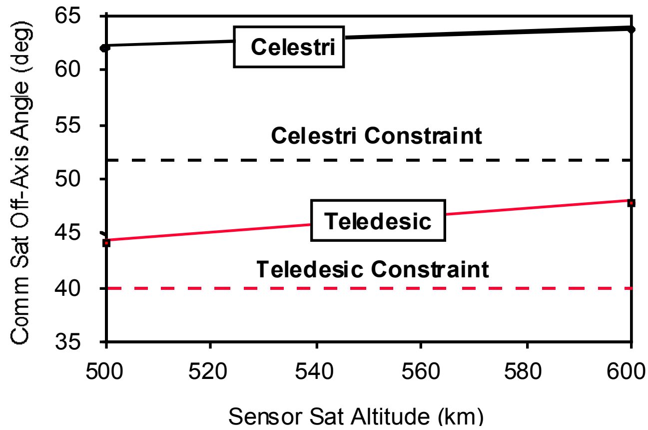


Figure 2. Maximum off-axis angle for relay link.

A similar question can be raised for the Teledesic constellation, which needs to link to the sensor satellite as much as 44 degrees off nadir. It is quoted as operating to a ground elevation of 40 degrees or an off nadir angle of 39 degrees. For the Teledesic arrangement, where the sensor satellites are in the same inclination as the Teledesic satellites, this problem would occur whenever the angle between the HSI satellite and the Teledesic satellite it is communicating with reached the maximum off-axis angle and there was no other Teledesic satellite available for handover (geometry to all Teledesic satellites outisde max off-axis angle of Teledesic). If the Teledesic satellite is truly limited to an off nadir angle of 39 degrees (e.g., actual performance may exceed specification), then an outage will occur approximately 17% of the time.

While both systems utilize phased-array antennas (electronically steered) and may therefore be physically capable of receiving signals outside the maximum off-axis angle, the commercial relay satellites may not be expecting a signal from outside the field of view of its earth fixed cells. Moreover, even though relative crossing velocities are less for the satellite-to-satellite link, they are different than what the current scheduling algorithms allow.

For a majority of the time the sensor satellite will be handed off to another relay satellite before the maximum off-axis angle is exceeded. However, there will be periods when the sensor satellite will be outside the maximum off-axis angle for all relay satellites, and a loss of connectivity will occur. Figure 3 shows this graphically for a Celestri constellation.

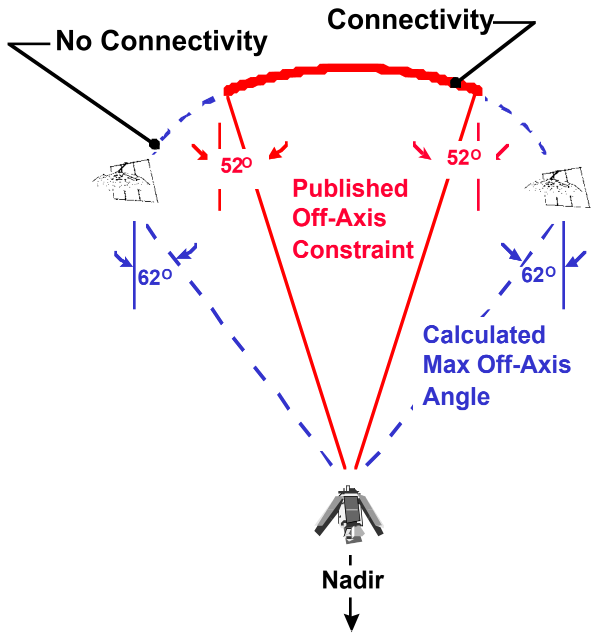


Figure 3. Gaps in Celestri relay link connectivity.

III. RESULTS. A qualitative comparison of the store-and-forward versus the commercial relay alternatives is shown in Table 4. A detailed summary of all results is shown in Table 5. Please note that the results of this architecture study are predicated on the study's ground rules; they may vary for different starting assumptions. In general, a broadband satellite relay may be a key mission enabler for a global remote sensing mission where an HSI payload (or an analagous high data rate system, e.g., space-based radar) intends to gather data over a significant portion of the orbit. Note, however, that other mission architectures exist whereby the satellite relay may provide little or no added benefit, such as with a military theater imaging system (Marshall, 1998).

a.) Cost. For the assumed collection frequency of 50 images per spacecraft per orbit, the life cycle costs for the 15-year mission of the satellite relay versus the store-and-forward concepts, at either 5-m or 10-m ground resolution, are comparable. For example, for 5-m resolution hyperspectral imagery, the predicted life cycle cost for using a Teledesic-type relay is slightly less than for the comparable store-and-forward system ($2.9 billion vs. $3.1 billion).

Table 4: Qualitative comparison of alternatives

Store-and-Forward	Commercial Relay
No outages,but limited by data storage capacity	Nearly unlimited data, but potential outage due to connectivity gap
Delay in receipt of data	Real-time return of data
Orbits can be tailored for coverage	Orbits potentially restricted to simplify communications interface

Table 5: Study results

	622 Mbps Data Rate		155 Mbps Data Rate	
	Sat Relay (Case 1)	Store/Fwd (Case 3)	Sat Relay (Case 2)	Store/Fwd (Case 4)
Constellation				
Orbit altitude (km)	500	500	500	500
Orbit inclination (deg)	88	83	64	83
Number of spacecraft	6	6	7	6
Number of planes	3	6	7	6
Revisit rate				
Maximum (min)	727	708	597	708
Average (min)	373	436	336	436
Spacecraft				
Payload				
Mass (kg)	198	198	73	73
Power (W)	68	68	22	22
Payload support				
Mass (kg)	34	31	20	13
Power (W)	110	173	110	49
Spacecraft bus				
Mass (kg)	262	310.9	173.6	119
Power (W)	426	603.2	385.4	211
Contingency mass (kg)	123	135	67	51
Propellant mass (kg)	65	79	47	31
Total S/C				
Wet mass (kg)	682	754	380	288
Power (W)	604	844	517	281
Deployment				
Launch vehicle	Athena 2	Athena 2	Athena 1	Athena 1
Number of S/C per LV	2	1	1	1
Launch margin	10%	49%	9%	21%
Replenishment				
Launch vehicle	Athena 2	Athena 2	Athena 1	Athena 1
Number of S/C per LV	1	1	1	1
Launch margin	53%	49%	9%	21%
Availability				
Number of life cycle S/C	25	25	30	25
Probablility of availability	96.7%	96.7%	96.8%	97.3%
Costs ($M)				
HSI satellite first unit cost	76	78	39	35
Total life cycle cost, based on 50 images per S/C per orbit	2,935	3,088	2,265	1,942
Max images per orbit in sunlight	Over 1,000	50	Over 1,000	50

The study did not assess comparable costs between the satellite relay cases and the store-and-forward cases for collection requirements greatly in excess of 50 images per spacecraft per orbit. However, the expectation is that as the data volume increases (either in terms of number of images or sensor throughput), the cost gap between the satellite relay case and the store-and-forward case will widen, with the satellite relay case being the more economical alternative. This is based on the prediction that the increase in lease costs of the satellite relay case will be smaller than the corresponding increase in store-and-forward satellite and ground station costs (due to larger on-board data storage requirements and increased throughput to the ground).

The potential for increasingly lower per-image costs for an HSI satellite relay system over a comparable store-and-forward system as image collection requirements increase beyond those assumed in this study, is a powerful concept that should be more fully explored. Sensitivity studies such as these are an important facet of architecture studies in addition to examining point designs.

b.) Surges in Collection Requirements. It is important to keep in mind that the on-board storage for each sensor satellite in the store-and-forward cases was sized for the 50 images per orbit. This capability is fixed at launch. The satellite relay cases, on the other hand, have an inherent flexibility in being able to dynamically vary their image collection requirements from the 50 images examined here to a thousand images or more per spacecraft per orbit. Admittedly for this type of high usage scenario, lease costs and contention with other users for the use of the commercial relay would increase and facilities and personnel for ground processing of collected images may need to be augmented. The point is that because the HSI spacecraft, under the satellite relay cases, are not constrained by on-board storage and ground station access to meet their image collection needs, surges in global imaging needs can be accommodated.

c.) Delays in Receipt of Data. The store-and-forward systems will experience a delay from acquisition of imagery to downlinking into the ground infrastructure, while the satellite relay concepts assume a real-time downlink of imagery through the commercial relay. For most imagery collection needs, barring military or contingency/disaster scenarios, this should not be an issue.

d.) Data Outages. The commercial relay system will potentially experience data outages due to gaps in connectivity when the sensor satellite is outside the maximum off-axis angle of the relay satellite. For Celestri, this can be as much as 68% of the time globally. While this study did not include data storage in the commercial relay cases in order to illustrate the differences between the two approaches, a prudent design would include some amount of on board storage to buffer data during connectivity gaps. On-board data storage may also open the possibility of buffering data from a payload design that takes data at a higher rate than the real-time satellite relay throughput allows (e.g., using the 5-m payload with the 155 Mbps relay satellites).

e.) Satellite Revisit. The current sensor definition uses a very small field of regard (± 30º), that is, the region on earth toward which the sensor can be pointed. For this reason it takes a large number of satellites to achieve a reasonable revisit time. The minimum elevation angle used by the sensor at 500 km is 57.4 degrees, which indicates that the sensor is limited to pointing near nadir. If the field of regard of the sensor could be increased, the number of satellites could be reduced significantly. Doubling the sensor swath size from ± 293 km to about ± 600 km would allow the maximum revisit time (MRT) to be cut in half (with a corresponding halving of the average revisit time ART). This doubling of the swath could be achieved by increasing the sensor field of regard from ± 30 degrees to ± 48 degrees. (Alternatively, an increase in HSI satellite altitude can be used to increase the swath width.) This nearly linear relationship between swath width and revisit time can be used to improve the revisit characteristics at the expense of a heavier sensor package. A trade study between the sensor weight and the field of regard might be useful to reduce the overall cost to achieve the desired revisit time.

IV. DESIGN DETAILS. Further details of the constellation, spacecraft and payload designs are provided in the following sections.

a.) Constellation Design. The hyperspectral sensor is to operate at an altitude of about 500 km with a field of regard of plus or minus 30 degrees as measured at the satellite. This means that the swath within which the sensor can look is only 293 km left or right of the ground trace. This small swath represents an earth central angle of only plus or minus 2.6 degrees. Equivalently, all ground points are viewed at an elevation angle of 57.4 degrees or greater. Using this sensor, it is desired to select a constellation of satellites to cover at least the region between 70º S and 70º N latitudes.

In order to provide coverage to the latitude band 70º S to 70º N, the inclination of the orbit must be at least 70 – 2.6 = 67.4 or more. In this study we adopted a lower limit of 64 degrees on the inclination of the orbits, which very nearly meets the goal.

The optimal constellations are shown Table 6. The column labeled "T/P/F" is the usual Walker notation for an optimized constellation. Note that in some cases Walker "star" patterns were found to be optimal. In these constellations, the values of RAAN (Right Ascension of Ascending Node) for the orbital planes are spread evenly over 180 degrees, instead of 360 degrees. These constellations are designated by an asterisk after the T/P/F value. The columns labeled MRT and ART are for the maximum and average revisit times at 34º N latitude. The goal was for a 12-hr (720-min) revisit time, which corresponds to a revisit of at least once during daylight.

In Case 1, the sensor satellites are at an altitude of 500 km and relay their data up to the Teledesic satellites at 1370 km. The Teledesic constellation consists of 12 planes, each containing 24 satellites. To make the relay link handover easier, each sensor satellite is located nearly in plane to a Teledesic plane. As it passes underneath the Teledesic satellites, the sensor satellite hands over from one Teledesic in the plane to the next. Since the Teledesic constellation's orbit is very nearly polar (approximately 88˚ inclined) the nodal regression rate caused by the oblateness of the earth is nearly zero.

Table 6: Optimal constellations

Case	Number of Satellites	T/P/F	Inclination (˚)	MRT (min) 34N	ART (min) 34N
1	6	6/3/2*	88	727	373
2	7	7/7/4	64	597	336
3,4	6	6/6/1	83	708	436

Further details of the constellation, spacecraft and payload designs are provided in the following sections.

The sensor satellite is also inclined at about 88 degrees to match the nodal regression of the Teledesic planes. For the constellation optimization in this case, the orbit altitude was fixed at 500 km and the orbit inclination was fixed at 88 degrees. The optimization program was free to place 6 sensor satellites in a subset of the 12 Teledesic orbit planes. For 6 sensor satellites, it chose a Walker star arrangement of 6/3/2*, that is, 3 orbital planes 60 degrees apart in RAAN, each containing two satellites. Using this arrangement it could achieve a max revisit time of 727 minutes in the region of 30 to 35 degrees north latitude. The average revisit time in this region is 373 minutes (see Figure 4).

In Case 2, the sensor satellites are at an altitude of 500 km and relay their data up to the Celestri satellites at 1400 km. The Celestri constellation consists of 7 planes, each containing 9 satellites. To make the relay link handover easier, each sensor satellite is located nearly in plane to a Celestri plane. As it passes underneath the Celestri satellites, the sensor satellite hands over from one Celestri in the plane to the next. Since the Celestri constellation orbit inclination is 48 degrees, the sensor satellite inclination must be chosen so its nodal regression matches that of the Celestri satellites. For the 500 km altitude of the hyperspectral satellite constellation, an inclination of 64 degrees is required to match the nodal regression of Celestri. For 7 sensor satel-

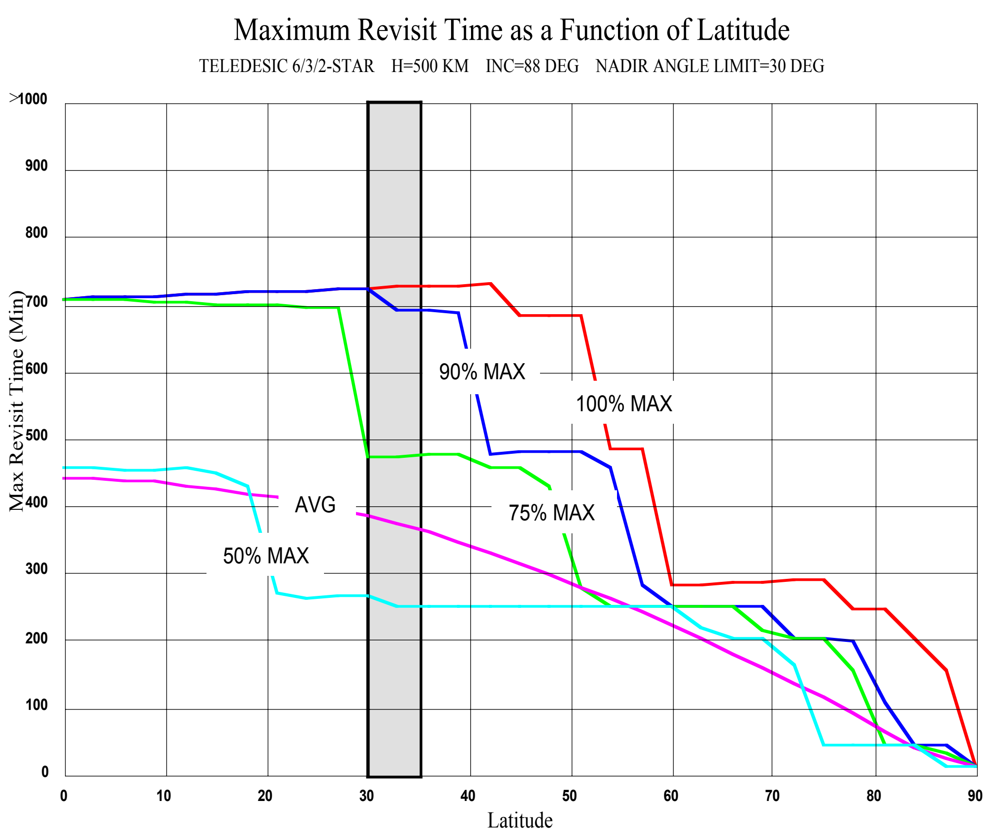


Figure 4. Constellation performance for Case 1 (Relay through Teledesic).

lites, the constellation optimization software chose a Walker arrangement of 7/7/4, that is, 7 orbital planes spaced evenly through 360 degrees of RAAN, each containing one satellite. Using this arrangement it could achieve a max revisit time of 597 minutes in the region of 30 to 35 degrees N latitude. The average revisit time in this region is 336 minutes.

Cases 3 and 4 are identical to the constellation optimization process. These cases are relatively unconstrained. While the altitude is fixed at 500 km, the other orbital elements are free to be optimized. The inclination, number of planes, and relative phasing are open for selection by the optimization process. Table 6 shows the T/P/F and inclination values which the optimization process selected for 6 sensor satellites. The corresponding values of MRT and ART are somewhat better than those of the more constrained Cases 1 and 2.

b.) Spacecraft Design. The spacecraft was sized to accommodate the payload in the desired orbit for the life of the spacecraft and to provide a means for transmitting payload data to the user. The spacecraft is divided into payload support equipment and spacecraft bus.

Payload Support Equipment. The payload support equipment consists of payload processing, payload data storage, and payload communications. It is sized based on the data rate and the storage requirements of the payload configuration as well as the type of communications architecture. Cases 1 and 2 require an uplink to the commercial communications systems with no data storage. Cases 3 and 4 require a downlink to the ground with data storage equivalent to 50 hyper-images over one orbit. Table 7 presents a summary of the payload support equipment for all cases.

In Case 1 and Case 2, it was assumed that the Teledesic and Celestri systems will be capable of handling a Ka-band uplink from space to space without disruption of services to ground users. It was further assumed that contact is continuous during data gathering/uplink and no data buffering is required. This method appears feasible in principle, but may not be operationally attractive from the viewpoint of Teledesic or Celestri.

In Case 3 and 4, it was assumed that contact is established with the ground station each orbit. Data storage was sized based on a calculation of sensor data volume assuming acquisition of 50 hyper-images per orbit.

Table 7: Payload support equipment characteristics

Case	Case 1	Case 2	Case 3	Case 4
Payload resolution (m)	5	10	5	10
Relay system	Teledesic	Celestri	Store-and-forward	Store-and-forward
Relay type	Uplink	Uplink	Downlink	Downlink
Communications band	Ka-band	Ka-band	Ka-band	X-band
Data rate (Mbps)	580	155	165	41
Storage (Gbits)	—	—	104	26

Spacecraft Bus. The spacecraft bus consists of the housekeeping subsystems of the spacecraft and includes propulsion, attitude determination and control (ADACS), telemetry, tracking and command (TT&C), command and data handling (C&DH), thermal, power, and structure. The spacecraft bus was sized based on the payload mass and power supported and the required Delta-V for orbit insertion and station-keeping. The spacecraft sizing algorithms represent a low-risk spacecraft using technologies representative of the technology freeze date of 2003. A growth contingency of 25% was added to the spacecraft mass.

In general, the launch vehicles were chosen to deliver this total spacecraft mass to the required orbit for the lowest cost. In Case 1 (Teledesic Relay) where there was more than one spacecraft per orbit plane, multiple manifesting on a single launch vehicle was possible for the deployment of the constellation. It was assumed that all replenishment launches delivered only one spacecraft to orbit A spacecraft mass summary for all cases can be seen Table 5.

In general, the largest spacecraft were required for the architectures using Teledesic for the data relay. The spacecraft mass was largely driven by the mass and power of the 5-m resolution sensor. The total spacecraft wet mass in this case was 682 kg. Spacecraft mass was considerably lower for the 10-m resolution sensor that was used for the Celestri relay. In this case, spacecraft wet mass was 380 kg.

c.) Payload Design. A hyperspectral sensor was chosen for this study because of the growing interest in hyperspectral imaging and because the high output data rates will stress a conventional store and forward data communication system. A sensor is categorized as hyperspectral based on the number of spectral bands the collected radiation is separated into. The sensor for this study uses 64 bands over the 0.4 to 2.5 micron wavelength region, stated previously. This fine spectral resolution permits discrimination among similar objects, such as species of plants, types of minerals and real versus decoy military targets. Proposed applications of this technology include early detection of diseases in crops and trees, location of valuable minerals, schools of commodity fish and pollution monitoring (Hardin, 1997).

The hyperspectral sensor design, as presented in Figure 5, uses a Cassegrain telescope to image an earth surface scene on the slit of the spectrometer system. At any instant in time, the sensor views a ground scene area defined by the projected image of the spectrometer slit. The sensor optical axis is directed at the satellite nadir point, and the slit image is oriented in the cross-track direction. In all cases in this study the pushbroom scan swath width was 8 km. A roll-axis gimbal allows the sensor optical axis to be rolled to plus or minus 30 degrees cross-track. The sensor is always operated in a purely pushbroom mode with a cross-track field-of-regard of plus or minus 30 degrees. Although the potential target revisit rate is increased as the maximum off-nadir angle increases, the performance degradation due to increased range, atmospheric path length and geometric distortion is probably unacceptable for angles larger than 30 degrees.

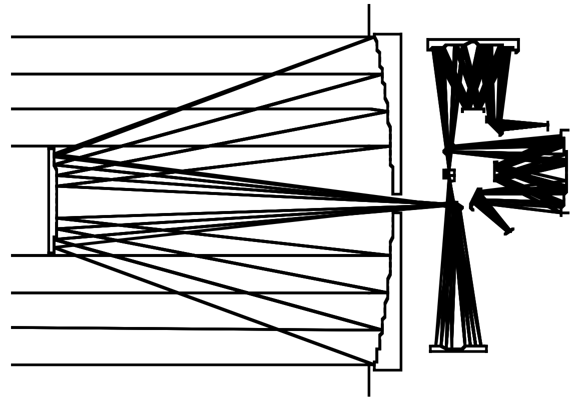


Figure 5. Hyperspectral sensor optics.

For this study, the baseline sensor design was the 90-cm aperture instrument operating at a 500-km altitude with a 5-m ground sample distance (GSD). This configuration produced an output data rate of 578 Mbps assuming 3:1 data compression and 12 bit analog-to-digital conversion. A single 8x8 km × 64-band hypercube is 655 Mb. This configuration was designed to fit within the 622 Mbps available in a Teledesic uplink. For this configuration, the estimated sensor signal/noise ratio (SNR) in the visible band is about 170, when viewing a solar-illuminated earth scene. No accommodation was made for backscanning the sensor line-of-sight to attain higher SNR ratios. To produce a data rate compatible with the 155 Mbps Celestri uplink, the GSD was increased while keeping the swath width constant at 8 km. As the GSD increases, the SNR also increases for a constant aperture size. For this study, we wanted the only sensor performance variable to be its spatial resolution, so the aperture size was adjusted to keep a constant SNR. Table 8 summarizes the key parameters of the two designs.

These point designs, combined with several previously developed designs, were used to develop equations used in the hyperspectral payload model. Curve-fits were done between the designs to develop equations that were incorporated into the model. This was anticipating that there would be some deviation from the two pre-defined cases. This capability was not needed for this study but now exists for future efforts.

All of the designs have gimbal-mounted telescopes. A 90-cm aperture telescope is quite large for a gimbal-mounted system, but the gimbal was used for consistency across the designs. One consequence of gimballing the 90-cm system is that it may not fit within the launch vehicle fairing. An alternative to a gimbal-mounted sensor would be to use satellite roll and pitch maneuvers for sensor line-of-sight pointing and scanning. However, in this study these satellite maneuvers were disallowed to avoid the requirement for compensating gimbal pointing and scanning for the communications antennas.

Table 8: HSI payload parameters

	Case Number	
System Parameters	**1,3**	**2,4**
Sensor Satellite Parameters		
Max output data rate (Mbps)	622	155
Satellite altitude (km)	500	500
Sensor Parameters		
Ground sample distance (m)	5	10
Aperture diameter (cm)	90	32
Effective focal length (cm)	200	100
Signal-to-noise ratio (visible band)	170	170
Size of 8×8 km hypercube (Mb)	655	164
Sensor mass (kg)	198	73
Sensor average power (W)	68	22
Sensor output data rate (Mbps)	578	145

Summary

This architecture study explored the trade space around a unique hyperspectral imaging system concept that utilizes a broadband satellite relay to downlink the sensor data. The broadband relay system offers the potential of virtually unlimited real-time distribution of data, at the expense of constraining the mission orbit to facilitate the interface to the satellite relay. Elements of mission concept, constellation design, spacecraft design and payload design were examined concurrently to derive several architectures that spanned satellite-to-satellite and store-and-forward cases. These cases were then compared in terms of cost-effectiveness (cost per number of images) and observations about the utility of using a link between an HSI (or comparable high data rate) payload and an LEO communications constellation.

The hyperspectral sensors described here generate data at a tremendous rate — one that will quickly fill any conceivable on-board storage device. A typical LEO satellite system might only be in view of a ground station for a few minutes out of each orbit. This combination of factors results in a restriction on the amount of data that can be collected by the sensor. By using a satellite relay for real-time downlinking of the data, the problems of on-board storage and ground station visibility are eliminated. This opens up the possibility of continuous data collection and provides a more efficient utilization of an expensive space-based asset. If the use of satellite relays allows the sensor to operate at a higher data rate, other performance improvements such as higher spatial and spectral resolution, additional wavelength bands or reduced reliance on data compression can be considered.

Acknowledgments. The hyperspectral imager application included in this paper is based on a study completed by the Aerospace Corporation for the Developmental Planning Directorate of the U.S. Air Force Space and Missile Systems Center (SMC/XR) under contract F04701-93-C-0094. Doug Holker of Aerospace's Strategic Planning and Awareness Directorate represented the customer. The study leads were Matthew Marshall and Dave Bearden. The following engineers participated in the study: Mark Barerra, Vince Canales, Evelyn Davalos, Andrew Dawdy, Shel Cantor, Marilyn Dubas, Amy Hagen, Bob Kellogg, Chris Kobel, Tom Lang, Carolyn Latta, Glenn Law, Kam Lee, Ted Milne, Jim Murrell and Bob Reid.

References

1. HARDIN, R.W., 1997, "Hyperspectral Imaging: How Much is Hype?" *Photonics Spectra,* Vol. 31, No. 7, pp. 82–94.
2. MARSHALL, M.F., N. LAO, J. NEFF, P. YUHAS, 1998, "Military Applications of Future Commercial Space Systems," AIAA Paper No. 98–5220.
3. Motorola Global Communications, Inc., 1997, *Application for Authority to Construct, Launch and Operate the Celestri Multimedia LEO System*.